U.S. DEPARTMENT OF COMMERCE
National Technical Information Service

PB-245 269

A GUIDE TO MODELS IN GOVERNMENTAL PLANNING AND OPERATIONS

MATHEMATICA, INCORPORATED

PREPARED FOR
ENVIRONMENTAL PROTECTION AGENCY

AUGUST 1974

KEEP UP TO DATE

Between the time you ordered this report—which is only one of the hundreds of thousands in the NTIS information collection available to you—and the time you are reading this message, several *new* reports relevant to your interests probably have entered the collection.

Subscribe to the **Weekly Government Abstracts** series that will bring you summaries of new reports *as soon as they are received by NTIS* from the originators of the research. The WGA's are an NTIS weekly newsletter service covering the most recent research findings in 25 areas of industrial, technological, and sociological interest—invaluable information for executives and professionals who must keep up to date.

The executive and professional information service provided by NTIS in the **Weekly Government Abstracts** newsletters will give you thorough and comprehensive coverage of government-conducted or sponsored research activities. And you'll get this important information within two weeks of the time it's released by originating agencies.

WGA newsletters are computer produced and electronically photocomposed to slash the time gap between the release of a report and its availability. You can learn about technical innovations immediately—and use them in the most meaningful and productive ways possible for your organization. Please request NTIS-PR-205/PCW for more information.

The weekly newsletter series will keep you current. But *learn what you have missed in the past* by ordering a computer **NTISearch** of all the research reports in your area of interest, dating as far back as 1964, if you wish. Please request NTIS-PR-186/PCN for more information.

WRITE: Managing Editor
5285 Port Royal Road
Springfield, VA 22161

Keep Up To Date With SRIM

SRIM (Selected Research in Microfiche) provides you with regular, automatic distribution of the complete texts of NTIS research reports *only* in the subject areas you select. SRIM covers almost all Government research reports by subject area and/or the originating Federal or local government agency. You may subscribe by any category or subcategory of our WGA **(Weekly Government Abstracts)** or **Government Reports Announcements and Index** categories, or to the reports issued by a particular agency such as the Department of Defense, Federal Energy Administration, or Environmental Protection Agency. Other options that will give you greater selectivity are available on request.

The cost of SRIM service is only 45¢ domestic (60¢ foreign) for each complete microfiched report. Your SRIM service begins as soon as your order is received and processed and you will receive biweekly shipments thereafter. If you wish, your service will be backdated to furnish you microfiche of reports issued earlier.

Because of contractual arrangements with several Special Technology Groups, not all NTIS reports are distributed in the SRIM program. You will receive a notice in your microfiche shipments identifying the exceptionally priced reports not available through SRIM.

A deposit account with NTIS is required before this service can be initiated. If you have specific questions concerning this service, please call (703) 451-1558, or write NTIS, attention SRIM Product Manager.

This information product distributed by

NTIS U.S. DEPARTMENT OF COMMERCE
National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22161

43F

PB 245 269

TECHNICAL REPORT DATA

(Please read Instructions on the reverse before completing)

1. REPORT NO.
EPA-600/5-74-008

2.

3. RECIPIENT'S ACCESSION NO.

4. TITLE AND SUBTITLE
A GUIDE TO MODELS IN GOVERNMENTAL PLANNING AND OPERATIONS

5. REPORT DATE
August 1974

6. PERFORMING ORGANIZATION CODE

7. AUTHOR(S)
Saul I. Gass
Roger L. Sisson

8. PERFORMING ORGANIZATION REPORT NO.

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Mathematica, Inc.
Bethesda, Maryland

10. PROGRAM ELEMENT NO.
1HA096

11. CONTRACT/GRANT NO.
68-01-0788

12. SPONSORING AGENCY NAME AND ADDRESS
U. S. Environmental Protection Agency
Office of Research and Development
Washington, DC 20460

13. TYPE OF REPORT AND PERIOD COVERED
Final

14. SPONSORING AGENCY CODE
EPA-ORD

15. SUPPLEMENTARY NOTES

16. ABSTRACT

This report is directed towards the majority of governmental (federal, state, and local) officials and their staffs who are confronted with the continuing task of selecting solutions, i.e., of making decisions in societal areas. The report is divided into 12 major chapters: 1. Introduction to Decision Models--a primer on the basics of models and modeling building. 2. Models and Policy Making, 3. Models in Air Pollution, 4. Models in Water Resources, 5. Models in Solid Waste Management, 6. Urban Development Models, 7. Models in Transportation, 8. Models in Law Enforcement and Criminal Justice, 9. Models in Educational Planning and Operations, 10. Models in the Field of Energy, 11. Models in Planning and Operating Health Services, 12. Econometric Models for Policy Analysis.

Each chapter is basically self-contained, with model concepts and terminology introduced in the primer. The Guide is an attempt to present to a non-technical governmental-oriented audience an overview of what models are, and how they have been used in a number of important social-urban areas. It represents a source from which governmental officials and others can obtain an understanding of this approach to decision making.

Reproduced by
NATIONAL TECHNICAL INFORMATION SERVICE
U.S. Department of Commerce
Springfield VA 22151

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
*Urban planning *Economic models Transportation Models *Decision Theory Econometrics Mathematical models *Government Forecasting	Solid waste Air pollution Social problems Decision models Government officials	05 K 12 A

PRICES SUBJECT TO CHANGE

18. DISTRIBUTION STATEMENT
Unlimited

19. SECURITY CLASS (This Report)
Unclassified

20. SECURITY CLASS (This page)
Unclassified

21. NO. OF PAGES
414

22. PRICE

281025

EPA-600/5-74-008

August, 1974

A GUIDE TO MODELS IN GOVERNMENTAL PLANNING AND OPERATIONS

prepared for

Office of Research and Development
ENVIRONMENTAL PROTECTION AGENCY
Washington, D.C. 20460
Contract No. 68-01-0788

EPA Review Notice

This report has been reviewed by the Office of Research and Development, EPA, and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the Environmental Protection Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

Abstract

This report "A Guide to Models in Governmental Planning and Operations" is directed towards the majority of governmental (Federal, State and local) officials and their staffs who are confronted with the continuing task of selecting solutions, i.e., of making decisions in societal areas. The report is divided into 12 major chapters: 1. Introduction to Decision Models—a primer on the basics of models and model building, 2. Models and Policy Making, 3. Models in Air Pollution, 4. Models in Water Resources, 5. Models in Solid Waste Management, 6. Urban Development Models, 7. Models in Transportation, 8. Models in Law Enforcement and Criminal Justice, 9. Models in Educational Planning and Operations, 10. Models in the Field of Energy, 11. Models in Planning and Operating Health Services, 12. Econometric Models for Policy Analysis.

Each chapter is basically self-contained, with model concepts and terminology introduced in the primer. The *Guide* is an attempt to present to a non-technical, governmental-oriented audience an overview of what models are, and how they have been used in a number of important social-urban areas. It represents a source from which government officials and others can obtain an understanding of this approach to decision making.

Contents

Chapter

	Preface	vii
1	Introduction to Decision Models	1
2	Models and Policy Making	39
3	Models in Air Pollution	61
4	Models in Water Resources	103
5	Models in Solid Waste Management	139
6	Models in Urban Development	165
7	Models in Transportation	201
8	Models in Law Enforcement and Criminal Justice	231
9	Models in Educational Planning and Operations	277
10	Models in the Field of Energy	317
11	Models in Planning and Operating Health Services	347
12	Economic Models for Policy Analysis	375

Preface

The *Guide* is directed towards a specific audience—the majority of governmental (Federal, State and local) officials and their staffs who are confronted with the continuing task of selecting solutions, i.e. of making decisions, in societal areas: If you are a part of our audience, we assume you have had limited scientific experience, but your elective or appointed tasks require you to resolve the conflicting objectives inherent in society's problems; problems which have scientific and technical implications, e.g. air pollution control systems, or which may be better solved by methodologies foreign to your experiences and training, e.g. the use of optimization procedures for routing refuse trucks.

From the realms of military, space and industrial problem solving, a body of knowledge, under the generic title of models, has been developed and has proven of great aid to decision making in those areas. This knowledge and its methodologies is now being adapted and applied to a broad range of domestic governmental problems. In these applications, some look at models and model building as an "aqua regia" which can dissolve all of society's problems; while to others, models are mysterious academic exercises and have not proven themselves in the arena of social action. This *Guide* is written from the perspective of a middle ground—to the extent that we understand the convoluted nature of these social problems and the needs of the decision makers, and to the extent that we have concomitant understanding of the power and limits of models, then models can aid the governmental decision maker to improve his ability to make more informed and effective decisions.

Our approach to imparting to you what we feel is an appropriate level and scope of information on models is to first present a primer (Chapter 1) on the basics of models and model building. This primer should supply the reader with a foundation for reading and understanding the remaining chapters which describe models in specific problem settings.

The primer is just that—it is tutorial, and not a handbook of models, nor a state-of-the-art in models; it is not exhaustive. In terms of our audience, the primer presents definitions, examples and discussions in a relatively nontechnical manner, but of sufficient strength so that the readers may not only understand the chapters which follow, but hopefully they will then be able to communicate better with model developers.

The succeeding Chapters 3–11 describe the use of models in a variety of social-urban fields: air pollution, water resources, solid waste, urban development, transportation, law enforcement and criminal justice, education, energy and health services. Each chapter is basically self-contained, but unless you have some knowledge of modeling, it requires prior reading of the primer. In addition, each author presents information on models which complement the material in the primer. Thus, for example, the section on "General Themes in Urban Modeling" of the "Urban Development Models" chapter describes modeling concerns which are applicable to most modeling situations; also, the chapter on "Models in Air Pollution" includes a technical discussion on stochastic models which is of a general nature. Most chapters have elements on model building which should reinforce the concepts and understand-

Preceding page blank

ing of readers in their pursuit of just how can models aid them within their decisionmaking framework. Thus a reader interested in models and, for example, air pollution, will not do justice to the total material if only the primer and the chapter on air pollution is read. Do read all, or at least read the nontechnical parts of all chapters so that each author's viewpoint and concerns are assimilated into a total viewpoint.

As econometric models are of general interest and are applicable to many areas, a somewhat more detailed description of econometric models for policy analysis is included as Chapter 12.

The fields selected for description in terms of the use of models were chosen for one of two reasons: (1) the field was well-developed and could contribute valuable lessons, or (2) the field is important even though use of models is embryonic and discussion here could foster wider use. Each author describes the critical decisions of his field, the models which have proven of value in analyzing some of the decisions, and specific applications. This approach is not adhered to strictly in that each author was allowed to define and delimit the scope of his field and to present his discussion based on his insights and understanding. Thus, in some instances, authors add their assessments and critiques if such discussions would add to the reader's knowledge and if such assessments could be made in a succinct fashion.

Some chapters are more mathematical than others. We felt that a set chapter format would be too restrictive for the authors and too limiting for the reader. A reader not interested in the mathematical presentations can skip those sections and still obtain valuable insights into models in governmental operations; the more technically inclined reader should allow more time to the pursuit of models, especially while reading the primer and the chapters on air pollution and econometric models.

In the structuring of a book for governmental decision makers, we must pay some attention to the scope of our definition of decision maker and decision problems. We sometimes interchange these concepts with policy makers and policy problems, or strategic vs. tactical, or planning vs. operations, high level vs. low level, etc. The primer sets the basic definitions for decision makers and their problems from a very general point of view. We find many applications of models to the decision problems described in the *Guide*. However, there is a concern that models are not being used to analyze significant policy problems at the highest levels of government, even though the application of models here can be of great value. The reasons for this lack of use and an approach to overcoming the deficiency are addressed in (Chapter 2) "Models and Policy Making."

In sum, the *Guide* is an attempt to present to a nontechnical, governmental-oriented audience an overview of what models are, and how they have been used in a number of important social-urban areas. It represents a source from which government officials and others can obtain an understanding of this approach to decision making.

SAUL I. GASS
ROGER L. SISSON

Acknowledgments

The following publishers have given their kind permission to use the cited materials:

From **AN INTRODUCTION TO MACROECONOMICS** by K. C. Kogiku, Copyright 1968 by McGraw-Hill Book Company, used with permission of McGraw-Hill Book Company, pp. 43–53.

From **THE ELEMENTS OF INPUT-OUTPUT ANALYSIS** by William H. Miernyk, Copyright 1957 by Northeastern University, Copyright 1965 by Random House, Inc., Reprinted by permission of Random House, Inc., p. 54.

Thanks are due to the many other authors whose reports on their research and applications of models served as the basis for most of the material in the *Guide*. The staffs of MATHEMATICA, Inc. (Bethesda, Maryland) and its subsidiary, Government Studies and Systems, Inc. (Philadelphia, Pennsylvania), were responsible for the creation, organization and editing of the *Guide*. Appreciation is also due to the following EPA personnel: Dr. Peter House, Director, Environmental Studies Division; Dr. Philip Patterson, Project Officer; and Mr. Gene Tyndall. Finally, sincere thanks to Ms. Carrie Scannell and Ms. Sally Whipp of MATHEMATICA, Inc. for their assistance in preparing the many versions of the manuscript to typed copy.

The Authors

INTRODUCTION TO DECISION MODELS

Roger L. Sisson is the Associate Director of Government Studies and Systems, Inc. He received his M.S. in Electrical Engineering from M.I.T. in 1950. His professional work is in design and use of planning systems for public institutions and in the development and application of models to support policy decision making.

MODELS AND POLICY PLANNING

Peter W. House is Director of the Environmental Studies Division, Washington Environmental Research Center, U. S. Environmental Protection Agency. He received his Ph.D. degree in Public Administration from Cornell University. For the past six years he has been involved in the design and construction of large-scale, policy-oriented models for the Federal government, including the CITY models and the RIVER BASIN MODEL.

Gene Tyndall is Chief of the Planning Coordination Division in the Office of the Secretary of Transportation. He received his B.A. in Economics from the University of Maryland and M.B.A. from George Washington University. He has over eight years experience in the design, development and application of management science techniques to policy-level problems, of which the past three have been within the Federal government.

MODELS IN AIR POLLUTION

Nozer D. Singpurwalla, Ph.D., is Associate Professor of Operations Research at George Washington University. He specializes in applied probability and statistics with particular emphasis on models for air pollution and reliability. He is co-author of a forthcoming book entitled *Statistical Methods for the Analysis of Reliability and Life Data*, John Wiley and Sons, Inc.

MODELS IN WATER RESOURCES

David Hunter Marks is Associate Professor of Civil Engineering at the Massachusetts Institute of Technology, where he is a member of the Water Resources Division and Director of the Civil Engineering Systems Laboratory. He holds a B.C.E. and M.S. from Cornell University and a Ph.D. in Environmental Engineering from Johns Hopkins University. His teaching and research is in both the water resources and systems areas; he is the author of numerous papers on water resource systems, water quality and environmental management and is co-author with Richard de Neufville of *System Planning and Design: Case Studies in Modeling, Optimization and Evaluation* to be published by Prentice Hall.

MODELS IN SOLID WASTE MANAGEMENT

Jon C. Liebman is Professor of Environmental Engineering at the University of Illinois at Urbana-Champaign. He received the B.S. degree in Civil Engineering from the University of Colorado, and the M.S. and Ph.D. from Cornell University. He is the author of a number of papers on modeling of solid waste collection, and has served as a consultant in this area.

MODELS IN URBAN DEVELOPMENT

JAMES C. OHLS is Assistant Professor of Economics and Public Affairs at Princeton University where he teaches courses in urban economics and in housing policy. He has an A.B. degree in economics from Harvard College and M.A. and Ph.D. degrees in economics from the University of Pennsylvania.

PETER HUTCHINSON is Assistant to the President's office, Irwin Management Company. He has an A.B. degree from Dartmouth College and a Master of Public Affairs degree from Princeton University.

MODELS IN TRANSPORTATION
and
MODELS IN THE FIELD OF ENERGY

KENNETH W. WEBB is a private consultant and a former staff member of the Urban Institute. He has an M.A. in Statistics from George Washington University. His research in transportation includes the study of network simulation with exploding queues and demand peaking. In addition, he has written on critical problems of transportation modeling with regard to social effects.

He has consulted with a major oil company on scheduling, manufacturing and related problems. In addition, he has designed a technology assessment approach to total energy studies.

MODELS IN LAW ENFORCEMENT AND CRIMINAL JUSTICE

SAUL I. GASS is a Vice President and Director of the MATHTECH Division, Washington, D. C. office of MATHEMATICA, Inc. He has a B.S. in education and M.A. in mathematics from Boston University, and a Ph.D. in Engineering Science—Operations Research, University of California, Berkeley. He was a member of the Science and Technology Task Force of the President's Commission on Law Enforcement and the Administration of Justice, and has conducted research efforts dealing with the application of operations research and computers to the law enforcement area.

MODELS IN EDUCATIONAL PLANNING AND OPERATIONS

EDMOND H. WEISS is a Senior Associate, Government Studies and Systems, Inc., in Philadelphia. He holds a Ph.D. from Temple University and an M.A. and B.A. from the University of Pennsylvania. Dr. Weiss is a consultant in planning and management systems for education, and is an elected member of the board of education in Willingboro, New Jersey.

MODELS IN PLANNING AND OPERATING HEALTH SERVICES

BOYD Z. PALMER is a Senior Operations Research Analyst at Government Studies and Systems, Inc. He has a B.A. in Mathematics and Physics from Earlham College and an M.S. in Operations Research from Case Institute of Technology. He has developed models for use in Comprehensive Health Planning, and has directed a project to design a simulation model of health services in an urban area.

ECONOMETRIC MODELS FOR POLICY ANALYSIS

E. PHILIP HOWREY is Professor of Economics at the University of Michigan. He earned an A.B. degree from Drake University and a Ph.D. from the University of North Carolina. His research and consulting activities have been in the area of national and regional econometric models.

Chapter 1

Introduction to Decision Models

By

Roger L. Sisson

SUMMARY 3

I. THE DECISION PROCESS AND THE ROLE OF MODELS 4
- Who are the Decision Makers? 5
- Styles of Decision Making 5
- Operational and Strategic Situations 5
- Decisions in Relation to Organizational Levels 6
- Scope of Decision Making 6
- Recurring or One-Shot Decisions 6
- The Nature of the Phenomena About Which Decisions are Made 6
- Concept of a Model 7
- The Basic Decision Process 9

II. TYPES AND CHARACTERISTICS OF MODELS 9
- Descriptive Models 10
- Physical Models 10
- Symbolic Models 11
- Simulation Models 11
- Use of Models 11
- Prediction and Optimization Models 13
- Deterministic and Probabilistic Models 13
- Dynamic, Semi-Dynamic and Recursion Models 14
- Continuity 14
- Feedback 14
- Degree of Aggregation 14
- Representing Behavior in the Model 15
- Some Standard Model Forms 15
- Simulation 16
- Econometric Models 19
- A Simple Econometric Model 20
- Input-Output Models 23
- Mathematical Programming 24
- Decision Trees 25

III. HOW MODELS ARE CREATED AND TESTED 26
- Recognizing the Need for a Model 26
- Problem Definition 28
- Choice of Analysis Method 28
- The Basic Steps in Model Design 29
- Model Specification 30
- Analysis of Data Requirements and Availability of Sources 31
- Creation of the Model 32
- Computation: Programming and Debugging 32
- Model Verification and Validation 32
- Model Use 32
- Analysis and Presentation of Results 33
- Implementation 33
- Documentation 33

IV. PROBLEMS IN USING MODELS 34
- Defining Measures of Effectiveness 34
- Aggregation 34
- Data Acquisition 34
- Verification and Validation 35
- Cost and Personnel Requirements 36
- Setting Up the Model 36
- Interpreting Results 36

REFERENCES 36

Introduction to Decision Models

SUMMARY

Most of this book is directed toward decision makers who have some information about models and wish to learn what kinds of models, relevant to their areas of interest, are available and what success has been obtained with them. We recognize that there are also a large number of public administrators who would like the basic information about models. The purpose of this chapter, therefore, is to provide a very brief, general background about models and how they help decision makers.

Modeling is a sophisticated combination of art and science. No one should expect to be able to understand it, even as an informed observer, without many weeks of study and actual practice in design and application. We do believe, however, that the flavor of modeling can be obtained by reading this chapter plus one or more of the subsequent chapters, those which are in areas of your interest. Every effort has been made to keep the subsequent chapters as free of jargon and mathematical notation as possible. This chapter uses technical symbolism only for one example.

In order to keep the total book within bounds it was not possible to develop extensive examples in this initial chapter. We hope the interested reader will read this chapter in conjunction with discussions in subsequent chapters which will serve as examples. Excellent texts which are available and are referenced throughout the chapter should also be used by the serious student in conjunction with this text.

This book deals with government decisions which are consciously made in complex and changing situations. The decisions usually are to allocate resources or to establish regulations or legislation. These are made by elected officials or by the heads of departments, i.e. by policy makers. We also deal with models to support some key decisions made by agency and bureau heads and program managers.

Models are computational procedures designed to help the decision maker and his staff predict the consequences of proposed alternatives. In some cases the model includes an optimizing procedure which computes the optimal alternative.

Section I of this chapter categorizes decision processes and develops the role of models in supporting decision making. Section II is a classification of types and characteristics of models presented in part to explain the terminology used by modelers. The discussion includes a description of the symbolism by which models are expressed and of the procedures by which they are used to predict or optimize. Several model types of most use in policy making are described.

It is important for decision makers to know how analysts develop and use models. In fact, for proper results, *it is necessary for the decision maker to interact with the analyst while the models are being developed and used* so that he fully understands the results. Section III describes how models are created, how data is obtained, how the models are tested and used in order to help compare alternatives and perhaps, equally important, to provide insights for the decision maker.

The final section of this chapter reviews some of the critical problems that can arise in the use of models.

If the reader has had no earlier contact with the concept of modeling, we hope that a thoughtful reading of this chapter will enable him to read at least the non-technical parts of the remaining chapters. We also hope it will encourage him to delve into technical areas of interest both within this book and to read some of the references.

I. THE DECISION PROCESS AND THE ROLE OF MODELS

The major theme of this *Guide* is the use of analytical aids in the process of making decisions. The making of a decision—the choosing among alternatives—can be viewed in many ways: as a psychological event, as a sociological drama, or as a process of data and information manipulation. If viewed in the latter way, the case we are considering, it becomes reasonable to consider mathematical, logical, computational and informational processing aids for improving the quality of the decision. We further assume that the decision maker is conscious of making a choice, although this is not always true. Among the mathematical-information aids, we shall be concerned with those termed *models* which are used to evaluate alternative decisions, and which can be applied to a wide range of important situations.

Where decision making is more than just impulse and guess work, it is preceded by some sort of analysis. If the analysis is made entirely by intuitive cogitation, then the concept of a model is not applicable. If, however, the analysis preceding the decision is an explicit activity, capable of being described, then models can improve the analysis and the results of the decision.

Decision making, especially where more than one person is involved, does not always occur at a sharply defined point in time. Some decisions evolve. Nevertheless, if any formal study precedes the period of decision, then a model may be of assistance in producing information for weighing the available alternative choices and probable result. Thus, this book deals with processes that involve the sequence of events: analysis, decision, action.

We illustrate the process for two decision situations which are of major concern to all levels of governmental planning and operations: *resource allocation* and *regulation setting*. In most decision situations an allocation of some resources must be made, Figure 1. The resources may be dollars (budget), manpower, facility or equipment capacity. The allocation may divide the available resources among a number of potential resource users such as agencies, programs, projects, services, departments, locations, client groups, or even individual clients. The allocation may also establish the resources to be available in various time periods.

A second type of decision, creating or changing regulations, often precedes and constrains the resource allocation problem, Figure 2. Examples of such decisions are: set acceptable pollutant levels or establish prices for scarce resources.

We next discuss some of the char-

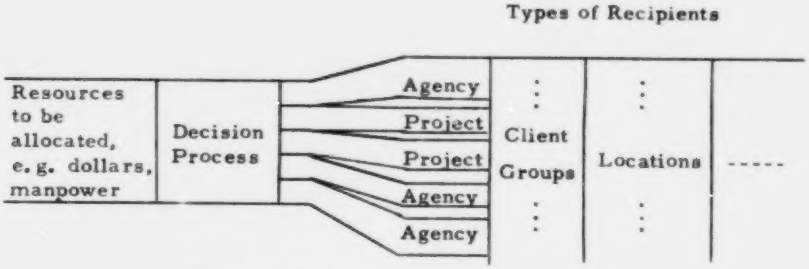


FIGURE 1—*Resource Allocation Decision Situation*

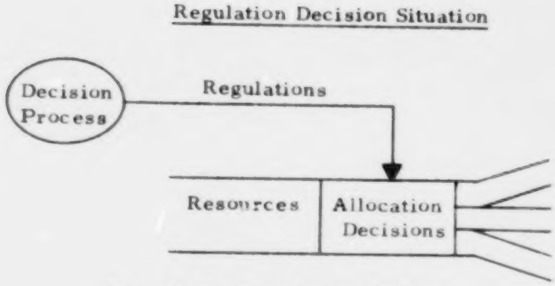


Figure 2—*Regulation Decision Situation*

acteristics or dimensions of decision makers and decisions.

Who are the Decision Makers?

In government the decision makers are usually well identified. They include elected officials, agency, department and bureau heads and program managers. Congress and other legislative bodies obviously make many important decisions, both regulations and allocations. Elected officials are sometimes distinguished from other decision makers by calling them policy makers. In the chapter on "Models and Policy Making," policy makers are defined as the group of persons who make the principal decisions which guide the public sector, and the process of policy making is defined as those actions taken by elected officials or their immediate staffs. The process of policy making at this highest level is differentiated from that at the department level by the scope of the problems and issues and the necessity to take into account the basic values of the whole society. This book emphasizes policy decisions.

Governmental decision makers may act individually, but more often than not they involve their staff and advisors. In some cases a task force has decision-making power, but usually the mission of a task force or a committee is to perform the analytic work preceding a decision.

Styles of Decision Making

Another dimension of the decision process is style, ranging from the least analytic to the most. Some decisions are made naively, even impulsively with no conscious thought involved. Decisions by experts or by good managers are presumed to involve some cogitation or judgment: some mental analysis.

Models are valuable in supporting those decisions which must be made in relation to complex and changing situations, where even informed judgment will not lead to the best possible results. Certainly, many situations facing government are of this nature. Models are important for many situations in government where it is either impossible or beyond fiscal feasibility to pilot test a range of ideas; that is, to perform experiments on proposed alternative programs to confirm which decision would be best. Nevertheless, where feasible, experiments are justified to provide a basis for decisions which have major, long-term effects on the clients of the jurisdiction involved. This is becoming an accepted approach in some social welfare areas and there have been experiments to evaluate alternative welfare, medical insurance, and housing subsidy plans.

Operational and Strategic Situations

A distinction is usually made between decision problems of an operational or tactical scope and those of a strategic or policy nature. Policy decisions almost always change relationships between other decision makers. Policy decisions directly or indirectly change who can decide what. The most useful way of recognizing the distinction between strategic and operational is by determining whether the actions which result from the decision can be reversed (or the effects of these actions compensated for) in the relatively short run, say within a year. The decision to assign an employee to a particular position within an office is a reversible, operational decision, since the employee can, without too much difficulty, be reassigned in the next month or the next year. On the other hand, consider a decision by a State to initiate a major rehabilitation service which requires the acquisition (long term lease or purchase) of a facility, the hiring of employees with at least implicit tenure and the raising of public expectations. This is obviously a long term, strategic commitment. It would be hard to reverse this decision within a month or even within several years of the date the program is initiated. It is important, of course, to make good decisions in either case.

Decisions in Relation to Organizational Levels

Generally, operational decisions are made at the lower levels and policy decisions at the higher levels of any organization. The right to make policy decisions is restricted in most organizations to certain top level positions through mechanisms such as budget approval, or limitations on the amount of money that can be committed on any one decision, or by the imposition of special review committee structures.

Decisions at lower levels tend to be prestructured by custom or by written procedures which are established by higher level decision makers to insure that routine decisions are made the way they want without having to review every one. Where a decision-making process can be completely defined, it can often be automated with the aid of modern computers.

Scope of Decision Making

Another dimension to the decision process is the scope of the implications of the decision. Some decisions involve only a specific person or work unit. A decision on how to locate offices within space assigned to a department has practically no consequences beyond that department. At the other extreme are all the global decisions made, say, by the U.S. about foreign commitments and activities. Scope is not always limited just because the level of government is low. For example, a decision made by a planning committee of a township in regard to sewage disposal can affect the environmental quality throughout a whole region.

There are also differences in "functional" scope, represented by the difference between "policy setting" and "planning." We have defined a policy decision as one that changes the structural relationships between decision makers. Obviously the creation of a new department or agency is a policy decision. The introduction of major new legislation which, by implication, causes changes or additions in organization is also a policy. The introduction of national health insurance, for example, would be a policy decision, because it would have significant effect on the organizational arrangements both in health insurance and in health delivery. Planning decisions, although having long-term consequences do not affect organizational interrelationships.

Recurring or One-Shot Decisions

Some decisions are made very rarely (yet usually are the most critical), such as decisions to institute a new program, to construct a major facility or to change a basic policy. Others are made annually such as to allocate funds through a budget. Some are made more often, such as to pursue particular alleged offenders of a pollution law. Since using a model may require a major initial investment in design, testing and data collection, it is sometimes hard to justify its use for a one-shot decision. The justification in these cases lies in the importance of the decision.

The Nature of the Phenomena About Which Decisions are Made

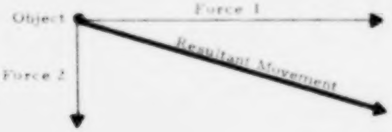
Because we understand the physical world much better than we understand human behavior, it is useful to distinguish decisions which have the effect of allocating physical resources from those which allocate human resources. It is also useful to distinguish those cases where human behavior significantly influences the decision from those where it does not. Most decisions made by government agencies are somehow influenced by human behavior, and this is one of the reasons why decision making and the use of supporting models is difficult in public administration.

An analogy might be appropriate here. In the physical world we know that if an object, Figure 3, is at rest and forces are applied to it in the direction and magnitude of the vertical and horizontal arrows, the resultant effect is that the object will move in the direction of the diagonal arrow. Our physical model predicts the result with certainty. However, if a sailor just in from a six month sea duty is the "object" and the forces are the attraction of two pretty girls in two directions, we cannot predict the resultant action of the sailor. We must approach the use of models for prediction of human behavior with caution!

With the above characterization of decisions in mind, we next turn to the nature of models and their relationship to the decision process.

Concept of a Model

A model is a way of abstracting the real world so that, not only the static picture of the world is obtained, but also the dynamic (time) interrelationships are represented. With an appropriate model of a real world situation, we should then be able to *predict* certain outcomes or determine how the real world would behave if we implemented a particular alternative decision. In some instances, our model will enable us to select the best or *optimum* decision. This book is concerned with models as they are used in decision-aiding analysis. Therefore, "model" never means "example" or exemplary case. In the literature, especially in the social fields, the word "model" is often used in this second meaning: as in a "model reading program," a "model environmental control policy." Such examples may be useful in defining a possible alternative, but are not models for prediction or optimization.



The models of use in informed decision-making tend to be *quantitative*. The interrelationships depicted by the model can be expressed in numerical terms, e.g. the relationship between a coal burning source and resultant pollution per ton of coal burned. However, appropriate methodology also exists for representing many *qualitative* factors.

Models have two major components: *variables* and *relationships*. In many real situations, it is possible to enumerate thousands of relationships and/or variables. The skill of the model builder enables him to capture the essence, only the important variables and relationships, so as to produce a meaningful and useful model. The creative input to the modeling process, the art of model building, is in translating a perception of the world into the essential relationships and variables, and thus into a model which is tractable and, hopefully, computationally manageable.

In the model, the *variables* represent either (1) values; numbers which are real world levels, counts, measurements, budget allocations, physical quantities, the results of survey instrument measurements, etc., or (2) codes which identify projects, items, units, people, groups, proposals, etc. The codes in turn can take on values such as zero if a project is not to be funded and a value of one if it is to be funded.

The *relationships* (equations, constraints) are expressed as procedures for computing the values of certain variables, given values of others: usually for computing the values of variables at a future time given the values at a present time. These computational processes represent the real interactions characterizing physical interconnections (the wear on a vehicle as it is used), economic relationships (the relationship between the demand for an item and its price), behavioral interactions (the effect of information produced by one group on the actions of another group) and many others.

A model represents some segment of reality. It is important to recognize the boundaries which define what is to be represented by the model. The included activities or structures are usually referred to as *the system* being analyzed. Sometimes these are further divided into related activities or *subsystems*. The boundaries may demarcate geographic or political limits, a temporal span, client groupings, funding arrangements, or behaviors or several of these. That which is outside the system is termed its *environment*.

In nearly all models it is found that the variables can be classified into four categories.

(1) *Controllable Variables*. These are variables whose values can be determined by the decision process. (For example, in a vehicle replacement problem, the factor we can determine or control is when to trade in the old unit for a new one.)

(2) *Uncontrolled Variables*. Many of the variables in a situation are not under the control of the decision maker (otherwise the problem would be trivial). Uncontrolled variables include physical phenomena (wear on a vehicle and, therefore, frequency of maintenance), variables which result from the actions of others or certain economic parameters, which are determined by the actions of the production and consumption processes in the country (e.g., price of a new vehicle).

(3) *Results or Output Variables*. These are variables characterizing the results of processes in the real world. (In the vehicle replacement situation a main output is the total cost of operation).

(4) *Utility or Value Variables*. The decision maker will set a utility or value on the results of the process.

Uncontrolled variables are of two types: those whose values are computed in the modeling process and those which are inputs to the model. The latter represent the effect of the environment on the system; the action of the world outside penetrating the system's boundaries. For the model to operate, it is clearly necessary to obtain estimates of the values of these input variables over the time span of interest. In some cases, these are known from existing data (such as the average cost of a particular kind of maintenance operation). In other cases, data about the future is required (such as the amount of trash that will be generated in a given area per week over the next five years). Estimating values of these uncontrollable variables may be done by other agencies, such as economic agencies or the Census Bureau. Or the analyst may use forecasting processes subsidiary to the model itself. Some modeling theorists believe that every model should be self-contained and should not require subsidiary forecasts or inputs. Such models, which include their environments, are termed *holistic* [13].

With any uncontrolled variable there is a certain degree of uncertainty: the estimates are never exact. Sometimes the estimates are sufficiently accurate and have a sufficiently small amount of variation so that we can assume they have a specific value. This leads to what are called *deterministic* models, the uncontrolled variables are assumed to be determined.

More often some of the uncontrolled variables are difficult to estimate or there is a variability which exists in the real world. (The amount of refuse generated in an area is not constant but varies from week to week.) In these circumstances, the most accurate model (not necessarily the most useful) will represent the variables as statistical quantities and take the variations specifically into account. These are known as *probabilistic* or *stochastic* models.

The utility or value variables are computed by a formula from the result variables; the formula is called the *objective-function*. Note that the assumption is that there is *one* measure of value. In some studies this value is measured in dollars, however, in many situations, particularly in governmental applications, the measure of value is more complex. Indeed, in many cases there is no single measure of the utility of the outcomes (this is termed a "multi-objective" situation). Whatever is chosen as an indication of value however, it is an important variable to be computed in the modeling process. The controllable variable values which give the best available utility are said to be the *optimum*.

The Basic Decision Process

In most cases the model builder assumes the decision problem to be of the following nature: *Find the values of the controllable (decision) variables which produce the greatest utility (value) as measured by the utility variable(s) given the assumptions about the uncontrolled variables.* For example, what is the best time to replace a vehicle to obtain the least total cost—the utility measure—given data about vehicle price and depreciation, maintenance requirements and cost and vehicle usage? The utility is computed from the result or output variables (costs, in the example) estimated by the modeling process.

Decision Theory is a title given to the general mathematical study of this class of problems. ***Operations Research*** is the study of specific applications of decision theory using mathematical tools. ***Management Science*** is the application of the results of decision theory and of operations research studies to specific management situations. (The terms management science and operations research are often used interchangeably, so that the definitions just given are not always clear cut unless the context is specified.)

Let us consider the model-builder in a practical situation. He must first identify the variables and then construct relationships (equations, computational processes) that interconnect the variables. In the most ideal case his work is facilitated by two sorts of theory: (1) theories about the physical and behavioral phenomenon in the situation, derived from physical, psychological, and sociological research; and (2) theories about the structure of the specific management process derived from operations research work. The theoretical results, if available, indicate to the model builder what are the most important variables, what are the relationships between them (in precise, computationally feasible form) and what procedures can be used to find the decision (controllable) values which will give the greatest utility.

All modeling processes allow us to compute present and future values of operating or result variables. Some allow us to compute the utility-optimizing controllable values. The latter is known as *optimization;* and (when considering situations over time) the former is called *prediction*.

II. TYPES AND CHARACTERISTICS OF MODELS*

The basic types of models available to an analyst can be classified by the manner in which they are expressed: descriptive, physical, symbolic and procedural. The latter two forms include those models which are more relevant to government planning and operations.

Any explanation of modeling methodology has two parts. The first part describes the form in which the model is to be expressed; the second part describes the way in which the model is to be used to make predictions (or optimize) and to aid in the decision-making process. Factors used to evaluate modeling methods include the relative cost of the use of the models, the ease with which the models may be communicated among technical people and from technical people to practitioners, and any special limitations on the application of the models. Table 1 is a taxonomy of model types based on these factors.

---

* This section adapted from Emshoff, Sisson, *Design and Use of Computer Simulation Models*, New York, Macmillan, 1970.

Descriptive Models

Descriptive models (which are expressed in native language) have many limitations, perhaps the greatest being that the method of prediction is usually internal and thus cannot be communicated nor easily replicated. The advantage of descriptive models is that the cost of making predictions is extremely low; therefore, they are the most commonly used model for decision making. The high powered, successful industrialist of our folklore probably had an intuitive sense for forming descriptive models and for manipulating the related information to yield effective decisions.

Physical Models

Physical models range from floor plan layouts to complicated aircraft wind-tunnel models. The method of optimizing with physical models is to search among alternative designs in the following way:

- Performance criteria are established.
- Estimates provide an initial combination of controllable decision variables.
- The model is then used to predict the value of the performance criteria under these conditions.
- A search rule, which incorporates all results to date, resets the controllable variables. The rule for search is designed to move the controllable variables in directions that will lead to the greatest improvement in performance.
- When maximum performance is found, the search has found an optimum, and the values of controllable variables that give this optimum are the desired operating conditions.

One significant advantage of the physical model is the ease with which its structure can be communicated to people with a nontechnical background. However, for modern decision-making purposes it suffers from an inability to represent information processes. Furthermore, there is often a high cost in the construction of a physical model, and, usually, the model can be used only for the particular problem for which it was designed.

Table 1. Taxonomy of Model Types

Model (Form of expression)	Method of Prediction	Method of Optimizing	Cost	Ease of Communication		Limitations
				Technical	Nontechnical	
Descriptive (Native language)	Judgment	?	Low	Poor	Poor (appears good but often misunderstood)	Cannot repeat the prediction process
Physical	Physical manipulation	Search	High	Good	Good	Cannot represent information processes
Symbolic	Mathematical	Mathematical	Low	Good	Poor	Needs previously developed mathematical structure
	Numerical approximation	Mathematical	Medium			
Procedural	Simulation	Search	High	Fair	Good	General properties not easily deduced from the model

Symbolic Models

Decision analysis has made major progress by the use of symbolic models. [5]. These models employ concise mathematical symbols to describe the status of variables in the system and to describe the way in which variables change and interact. Predictions or optimal solutions are obtained from these symbolic representations by means of mathematical, computational and logical procedures, called deductive methods. The cost of using symbolic models is often low as most models of this type can be solved by standard computational procedures. The conciseness inherent in symbolic models makes them particularly suitable for communication among technical people. On the other hand, the symbolism is not familiar to the average decision maker, and so communication between the technical person and the nontechnical person is difficult. (This situation is improving, however, as more people are becoming trained in the use of mathematics and in symbolic and computer languages.)

Simulation Models

The fourth model type identified in Table 1 is procedural; however, it is generally referred to as simulation. The model is actually a procedure expressed in precise symbols; the term simulation refers to the method by which the model is used to make predictions. In a sense, every model is a simulation of reality, but in this book, in consonance with current modeling usage, the term is restricted to mean the procedural model and its execution process.

By means of a series of elementary operations on the appropriate variables, a procedural model expresses the dynamic relationships hypothesized to exist in the real situation. These operations are usually stated in a computer-like flow chart, but may also be expressed in the form of decision tables or other procedural languages. The prediction of outcomes is made by actually executing the procedural steps with appropriate initial data and parameters. The execution process may be carried out clerically, but it is generally done using a computer. The computer is programmed to perform the procedure (which is the model); running the program creates the prediction. Concisely, then, a simulation is a model of some situation in which the elements of the situation are represented by arithmetic and logical relationships and processes that can be manipulated (on a computer) to predict the dynamic properties of the situation.

In contrast to symbolic models, the problems usually attacked by simulation procedures do not lend themselves to solution by standard computational techniques. Thus for simulation as with other model forms that do not utilize an explicit mathematical calculus, appropriate combinations (solutions) of controllable variables must be found by a search process or some other form of enumeration. Simulation is expensive in terms of both the model preparation time (although this has been reduced significantly by use of simulation languages) and the cost of the computer time necessary to make the predictions during the search for the optimum. An important by-product of a simulation model is that, properly written, it facilitates explanation to people with a nontechnical background. Simulation models can be created for almost any situation in which a researcher can state precisely the relationships among variables, although it may be difficult to obtain the data for such models.

This book is about symbolic mathematical and simulation models and their use in a wide range of governmental decision-making settings.

Use of Models

Models may serve different purposes in relation to specific analyses. At least the model should be a predictor to help the analyst predict the values of the output variables (given the assumed uncontrollable input variables and specific controllable or decision variables). The model is used to answer "what if" questions of two types: (1) "what if we set the decision variables at certain values?

(e.g., 'what if we traded the vehicle every five years')?" and (2) "what if an uncontrollable takes on a different value?" (e.g. 'what if maintenance rates increase at 10% per year instead of 6%?'). The total decision problem is, of course, to find the optimizing values of the decision variables. If the analyst has only a predictive model, then he must use the predictor to search for the satisfactory values of the controllable variables as described above. It is never possible to try out all possible combinations of values of the controlled variables. There are just too many in any real situation. One can try a large number at random and pick the one predicted to have the highest utility, but there are more organized ways for searching for best values [11].

An optimizing procedure has two characteristics: (1) it is a well defined series of steps which, for real problems, can be executed in a reasonable amount of computational time; (2) there is a mathematical proof which guarantees that the optimum solution is found at the end of the computational process. There are some intermediate analytic tools where the procedures do not meet criteria (2), but can be shown to yield good, nonoptimum values. These are called *heuristic* procedures in that they involve special logical, mathematical and sometimes intuitive steps which are based on the structure of a particular problem and which have not been shown to lead to an optimum solution.

Table 2 is a brief summary of some decision situations for which models have been developed. It is not meant to be all inclusive. Most of the standard models and their uses are described in other chapters. An informed decision maker should know enough about modeling to recognize a situation for which a model might exist. Models most relevant to policy making are described at the end of this section. The informed manager should also understand the deductive methods used to compute model predictions, particularly calculus [14]

Table 2. Summary of Types of Decisions For which Models Have Been Developed

Situation	Decision	Organizational Level of User	Model Name
Queuing, waiting for service	How much service capacity and how organized, priority rules	Low	Queuing, Sequencing, Simulation
Facilities that wear out	Timing of maintenance and repair	Low	Replacement
Items stored for future use	Amount and timing of orders	Low	Inventory, Simulation
Allocation of resources to activities	How much of each resource to allocate to each activity	Medium	Mathematical Programming, (Linear, Nonlinear, Goal), Dynamic Programming
Market place distribution of goods, money or services	Price constraints, market regulations, subsidy levels, etc.	High	Econometrics, Input-Output, Industrial Dynamics, Simulation, Markov
Determine whether to obtain more information before a decision (by survey or experiment)	Amount to spend on information acquisition	High	Decision Trees, Bayesian Analysis
Competition for limited resources	Strategy to employ	High	Game Theory

and the statistical methods used to estimate factor values [15, 16]. Typical decision problems which exemplify the use of models are:

- *A local sanitation agency operates* vehicles to collect refuse. One operational decision it must make is when to replace a vehicle. If the vehicle is allowed to get too old, maintenance costs become high. But there is a real cost in purchasing a new unit in terms of both the purchasing process and early, large depreciation in value. Thus there should be some best or "optimum" period after which the vehicle should be replaced. This is a relatively well-defined operational decision for which models are routinely used. [1].
- *A metropolitan air conservation commission* has the problem of setting air pollution regulations affecting the sulfur content of coal burned in the area. A "linear-programming" model was used to evaluate the economic feasibility of controls on coal by examining all sources of coal pollutants in the area's airshed. [2].
- *A Federal Agency* must consider the question of setting the ceiling price of natural gas. This is a problem in economics and use of an "econometric" model can be of value. [3].

Models vary greatly in terms of the way they are constructed and the way they operate. We next define a series of model characteristics. Knowledge of these can help one to understand or select a proposed model for a particular application. These characteristics are used to describe the models in the other chapters of this book.

Prediction and Optimization Models

As we have pointed out the *ideal model* has two aspects or parts, the first part represents the real world and allows us to predict how that world might unfold (given certain assumptions about both uncontrollable inputs and about the decision or controllable variables). The second part then selects from the feasible range of controllable variable values the particular ones that optimize or give us the most desirable results. (The two parts are not always clearly separate.) Some models accomplish both aspects and are called optimizing models. Other models (because of limitations on the state of the art or the skill of the modeler) contain only the predictive process. Usually a study of the outputs provided by the model will indicate quickly whether it makes a prediction or finds an optimum. As some optimization models, as well as most other types of models, cannot always represent institutional and organizational constraints in an explicit manner, these type of constraints must be considered when attempting to apply a particular solution.

Deterministic and Probabilistic Models

In deterministic models it is assumed that the exact values of all variables can be computed and values of all parameters are known. In a probabilistic model, at least some variables or parameters have an unpredictable randomness and must be represented as statistical variables. If the model is dynamic it might be termed *stochastic*, i.e., the model includes random variables that depend on some parameters, often a time variable.

Probabilistic quantities, although not known with certainty, are not complete mysteries. What we do know about them is usually characterized by a *distribution*. This tells us how often the probabilistic quantity is likely to have a certain value. A simple distribution is shown in Fig. 4.

This variable has a value of four 50% of the time, three 40% and five 10%

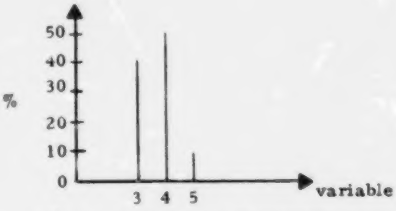


FIGURE 4—*Frequency Distribution*

of the time. Distributions are characterized by measures like the *mean* or average; and by measures of the spread of the distribution about the mean like the *variance* or the *standard deviation*.

Dynamic, Semi-Dynamic and Recursion Models

Most decision-making situations require a model which represents situations which change over time. Thus most models are dynamic in the sense that time is explicitly represented and the variables change with simulated time as they would (if the model is correct) in the real world. Most such dynamic models are completely self-contained, i.e. the model and other variables change from time period to time period automatically. There are a few models however, which are only semi-dynamic. That is the model computes the transition from one time period to the next, but then the analyst may inspect the output and adjust certain inputs and parameters within the model before the next time period is computed. This is done largely with the large, more complex, econometric models designed to represent the economics of an entire nation. The complexity occurs largely because the model has to represent the various markets. The analyst introduces factors into the model which are not automatically modeled: factors having to do with legislative and executive fiscal and regulatory actions, strikes, or other relatively unpredictable events. The official designation of these semi-dynamic models is *quasi-equilibrium* models.

Continuity

Another characteristic of models, closely related to the level of aggregation, is the continuity of the variables. Continuity, simply stated, means that the variables move smoothly from one value to another and do not jump around, i.e. they can take on all fractional values as well as integers. Many elements which are actually discrete, that is, can take on only unit values like 1,2,3, , can be represented as continuous. For example, the population of the nation actually consists of whole people but for national planning purposes can be considered as an aggregated continuous variable since we are not interested in the exact value. Many factors, however, cannot be represented as continuous. For example, in assigning individuals to jobs either the assignment is made or not. The discontinuity must be incorporated into the model.

Feedback

Feedback is a property of a model and presumably of the situation it represents. Suppose a situation involves a cause-effect chain:

$$\begin{tikzcd} A \arrow[r] & B \arrow[r] & C \arrow[r] & D \end{tikzcd}$$

but B is determined not only by A but also by D

$$\begin{tikzcd} A \arrow[r] & B \arrow[r] & C \arrow[r] & D \arrow[ll, bend left] \end{tikzcd}$$

The D.B path is the feedback loop. For example, suppose consumer attitudes cause business levels which cause tax increases, but tax increases cause changes in consumer attitudes. We then have feedback. If the feedback *adds* to the earlier effect it is *positive feedback* and often causes oscillations or cyclical changes in activity. If it subtracts it is *negative feedback*, and tends to increase the system stability. *Control theory* is the study of feedback systems, [13].

Degree of Aggregation

An important characteristic of a model is the level to which it aggregates real world phenomena. Obviously this characteristic can be evaluated only in relation to the kinds of situations to which the model applies. For example, let us consider models intended to assist in making decisions about welfare allocations. The most detailed model would represent every individual who could come into the welfare process. (In order to represent the future, the model would have to simulate the birth and life of people as yet unborn.) The most aggregate model would represent the entire body of welfare recipients in the nation as a single entity; perhaps by one variable such as their income level (as is done in some econometric models). In between is a range of model structures which are more aggregate than the individual level but less than single representation. A welfare model may be disaggregate in terms of geographic areas, in terms of social-economic characteristics, in terms of current welfare status, in terms of basic need, in terms of age groups, etc. Obviously the more aggregate models are simpler in that they have fewer variables and few interrelationships. Thus aggregate models are less expensive to use: they require a less extensive data collection and preprocessing, and they operate quickly when computerized. On the other hand, aggregate models stand a greater chance of being unrealistic for they deal with artificial groupings and not with direct representations of reality. In fact, a good deal of the art of modeling is in how to aggregate so as to gain the benefits of an efficient model and yet retain useful and realistic representations.

Obviously the selection of the level of detail in the model also depends upon the nature of the decision. If we want to make decisions about the allocation of funds to specific agencies, we must, of course, use a detailed model.

Representing Behavior in the Model

Some of the relationships in a model may represent the behavior of people, e.g., the way people decide what job to take, the way in which a student learns, the decision-making behavior of a foreign power, or the choice of transportation mode. In order to obtain a mathematical or simulation model of such behavior it is necessary to have a basis of understanding the behavior. Often such understanding does not exist, empirically or theoretically.

In order to perform research on the behavioral situation, the analyst tries to study it in the real world by observation and surveys. But often this is expensive or infeasible. The analyst might then try to use a partial simulation in which the behavior is acted out by a real person, but the rest of the situation is simulated. When humans are used as elements of a modeling process, the model is called a *game*.

In a few cases the game may involve the actual persons. For example, the analyst may have welfare recipients making decisions. More often however, he chooses a surrogate (a college student for example) to play the *role* of the person being modeled.

Thus, in a game model many of the aspects of reality are modeled by computations; only certain of the behavioral relationships are modeled by the human. Game models are rarely useful in support of actual decision making because they involve too many unknown assumptions and uncertainties. Their main use is in training, research, or for communication purposes.

The word *gaming* usually means the use of game models as defined above. Sometimes, however, it refers to the use of a predictive model to try out (search for) different alternatives. In contrast, *theory of games* is a particular mathematical theory which can be applied to situations in which there are two or more parties competing for resources. To date there has been limited application of the theory of games in the public area.

Some Standard Model Forms

A manager should be acquainted with the more common forms of models, just as he should be acquainted with the techniques of personnel management, leadership or any other phase of his trade. In this section we will introduce in as easy way as possible the most common policy-decision-supporting models, i.e. those of simulation, econometric, input-output, mathematical programming and decision trees. There are several texts [5, 6, 1] that contain simple explanations of models for operational decisions.

Although some of the discussion which follows next is of a technical nature, it is felt that a close reading of these model examples will enable the reader to obtain a better understanding of the material presented in this book.

As we have mentioned, a model can have two parts: (1) the part that represents the real world processes and allows the examination of the effects of changes in them and (2) the process for optimizing. Policy situations are very complex and for them only part (1), a representation that facilitates prediction, is often all that is possible. We do not yet know how to optimize in any formal way for most situations. Let us look at the policy decisions and how the models relate to them; see Table 3.

The four kinds of decisions shown in Table 3 fall into two basic groups: (1) the establishment of a new program or constraint (including forecasting to support such decisions) and (2) the allocation of resources to existing programs. The latter decisions tend to be much more specific, because the nature of the programs are already defined and their consequences are often clear either from past experience or by logic and deduction. In allocation there are many complications, and we do not wish to imply that allocation of resources is simple, but it is less difficult than the other group.

Most policy-supporting modeling techniques fall into the general class of simulation. The other model forms, such as the econometric, differ from general simulation only in that they employ one or more theoretical relationships and, therefore, have a more definite structure. The theory generally deals with assumptions about how groups of people will behave under specified circumstances. Perhaps the most well-known of these is the supply-demand relationship of economics, which says that in a market situation, as a product or service becomes scarce, people will bid for it so as to increase its price (and conversely).

In the case of model forms relevant to policy decisions it is almost always the case that the method of producing the prediction is to execute on a computer the relationships made explicit by the model. That is, the computer is used to calculate the model relationships which in turn simulate (as best the modeler can arrange it) the real process. In the next few pages we will describe briefly each of the important model forms and try to show:

- — the basic assumptions or theoretically derived structure,
- — how the model is constructed,
- — how the parameters of the model, that is, the constants, are estimated,
- — how assumed programs, allocations or changes are input into the model (to make it represent a particular alternative to be studied),
- — how the model is used or computed in order to make a prediction of what would happen given the alternative.

Simulation

Simulation (which was described above) is actually a general technique for creating models and, in a sense, is more general than any particular model form [4]. In simulation the analyst programs a computer directly to represent the situation under study. The only limitation to the situations that can be modeled are those which the analyst imposes upon himself because he lacks an understanding of the relationships within the real situation. In addition to most physical, chemical and biological processes, human behavior can be modeled if the analyst dares hypothesize how the behavior of people can be represented. Researchers have modeled such phenomena as:

- — the results of election campaigns,
- — the effect of advertising and pricing of pharmaceuticals on the physicians' behavior in buying them,
- — an urban area including the major functions such as industrial growth, residential growth, transportation, commerce and pollution,
- — the whole world (of course, at an aggregate level).

Many of the models discussed in other chapters of this book are simulations.

The essence of simulation, and indeed most of these modeled forms discussed in this section, is the following: the analyst can understand by observa-

Table 3. Main Features of Policy-Supporting Models

Decision (or Study)	Example	Model Forms	Method of Estimating Parameters	Basic Deductive Theory	Method of Computing Prediction
Establishment of incentives, constraints or regulations	Impose (or remove a) limit on the price that can be charged for natural gas (at wellhead)	Econometric	Statistical methods	Economic supply demand & simultaneous equations & calculus	Computer
		Input-Output	Statistical methods	Matrix algebra	Computer (manual in simple cases)
Approval and initiation of programs or projects	Pass legislation for a health insurance subsidy to people	Simulation	May be simple "guesstimates" or formal statistical methods	(none)	Computer
Allocation of resources to programs	Appropriate funds for subsidy programs: agriculture, housing, health food	Mathematical programming—linear	May be informal or formal	"Simplex" method	Computer
		Mathematical programming—nonlinear		Various special methods	Computer
		Simulation		(none)	Computer
Forecast overall trends to pinpoint problems & suggest possible solutions	Forecast pollution levels in a region	Simulation	Usually formal statistical methods	(none)	Computer

tion, analysis or theory the relationships between two or three factors. That is, he can understand small bits of the total process. He, or the decision-maker or anyone else has a difficult time, however, understanding the whole process and making predictions about it. In the simulation process these understandings of the small parts of the system are assembled in a precise way into a complete model. The model then is more likely to make a good overall prediction because it is based on a series of small, good understandings rather than one large, poor one.

In the most basic form of simulation time is represented as discrete points (not a continuous process). The world is viewed as a series of snapshots, with the program computing the changes between snapshots. The time interval can be any size (minute, year, decade) or even unequal, e.g. to the next significant event. The computations can represent any change large or small. In one form of simulation, called *continuous*, time is always advanced in equal steps, small compared to the phenomena under study and the computations assume that all variables change by a small amount each period, an amount determined by explicit rate-of-change variables (which themselves can change a little each time interval). Continuous models are often useful in studying long-term aggregate system behavior.

One of the most difficult aspects of a simulation model is to determine its validity. Since the model may be prepared on the basis of one (or a small group of) analyst's understanding of the important relationships, its validity may be more uncertain than models which are based on more widely understood and accepted relationships (that is, on proven theory).

Because of the generality of simulation there are no basic assumptions inherent in the technique. The assumptions are those which the analyst incorporates and, therefore, it is very important for any user of simulation results to be fully briefed on those assumptions.

The construction of simulations follows the process to be discussed in Section III. The analyst first studies the situation and gets firm in his own mind the variables and relationships involved. He then expresses these in a form which facilitates the preparation of a computer program. One such representation is the computer flow chart: a series of boxes which describe the computations which represent the relationships in the model. An aid to preparing the computer program is a simulation language: a language, in the usual sense, with verbs and nouns and rules of grammar. A language is designed to help the analyst describe a certain class of situations. The most common simulation languages are:

- **DYNAMO** [13], for preparing continuous simulation models of socio-economic systems,
- **CSMP** [4], for preparing continuous models; originally intended for representing physical (mechanical, chemical, electrical) systems, but useful for some economic models,
- **GPSS** [4], for discrete models intended for representing operational situations (e.g. traffic),
- **SIMSCRIPT** [4], actually a language to aid the analyst; not really a user-oriented language.

A user of a simulation model should understand the essence of the simulation language, if used, so that he can communicate with the analyst about the assumptions in the model. Most of these languages have been designed to be relatively easily understood because their purpose is to describe the world from the decisionmaker's viewpoint, not from the programmer's.

Any of the available techniques may be used to estimate the values of the parameters in a simulation model. In the complex situations which policymakers face, it is usually difficult to use any of the formal methods of estimation. Either there is not enough historical data available, it is not in the right form, it is too expensive to obtain, or the agencies which have the data cannot make it available. Under these circumstances the art of guessing at the values of the parameters should not be belittled. Guessing at a series of parameters and then using these in a well thought-out model to make a prediction probably leads to better predictions (and ones which are easier to understand and test) than simply guessing at the overall prediction itself.

Alternative cases are set up by redesigning that portion of the simulation which represents the changes that would result from the proposed alternative. This can be a time-consuming task, but is facilitated if the model is expressed on a simulation language. When the simulation program has been designed a computer program is prepared to carry out the procedures. This may be done by the analyst or a programmer. Where the model has been prepared in a simulation language, the translation to the computer program is automatic. The program is tested: the data representing the parameters and the initial state is fed into the computer; the computer program is executed for as many time periods as desired. The output of the computer is the desired prediction. The simulation may be re-run a large number of times in order (1) to get a better statistical representation of the prediction and (2) to examine a number of different alternatives.

Econometric Models

Econometric models are, in a sense, simulations in which most of the key variables and most of the relationships between them are derived from economic theory [23]. The other difference is that econometricians usually insist that the parameters in the model be derived from carefully designed statistical estimating procedures which use past economic data as a basis. This fact, on one hand, tends to increase the model's validity, but, on the other, limits the scope, since economic historical data is readily available only for certain aggregate phenomena. Econometric analysts also are willing to make adjustments to the model between each year simulated on the basis of special information (e.g. predicted legislative changes); so that there is some human "tuning" of the model as it computes predictions.

A *market* is a situation in which prices, established by many individual transactions, cause supply and demand to come into useful relationship. There are markets in consumer goods, industrial capital goods, investment funds (with interest rates being the price) and in labor (with wage rates as the price).

Many State and Federal governmental policies, as expressed in legislative acts, executive orders or commission rulings, are designed to alter or constrain relationships in a market situation. Thus, models which can represent and predict the consequences in such situations are particularly valuable in governmental planning. Econometric models fill this role.

> "Economists construct . . . models to study the behavior of various economic units in the activities of producing, exchanging, and consuming economic goods. Economic units are households, firms, governments, etc. Macroeconomic models are economic models that concentrate on various groups of economic units using aggregative measures such as national income, total employment, and the price level. Each of these economic units may have definite motives such as maximizing satisfaction or maximizing profits. . . . Macroeconomic models examine how the interactions of various sectors with different behavior patterns determine the aggregative magnitudes of the total economy. Some macroeconomic models go even further and explore the possibilities of influencing the aggregates to attain desired goals by public policy actions." [23, p. 1]

The basis for the development of an econometric model is assumptions about how economic units behave.

> "There are many simple theories based on assumed motives of various groups within the economy. After all, assumptions come cheap. If we know exactly what each

> group's motives are, then we can fairly safely deduce the consequence of interactions of these forces. But not all motives are apparent. For prediction we must depend on observed behavior. A case in point is the controversy concerning the consumption function, which over the years has produced a variety of explanations" [23].

Consumption can be assumed proportional to absolute income, income relative to others, preceived long term income, etc. Many econometric models have evolved over the past forty years. At first they were largely used for theoretical development. Since about 1945, they have been used for predictions in support of policy-making, although, for most applications they are in an advanced research or developmental stage. Early models represented a nation as a whole and were useful only for overall Federal policy analysis. Current models contain tens or hundreds of "sectors." A sector is an identifiable part of the economy: an industry, a governmental function, a group of households. Thus modern models should be useful for analysis of rather specific proposals.

Most models are for analyzing long-term consequences of policy changes. The scope of a model in regard to the sectors of activity covered depends upon the purpose of the analysis. Models to assist in studying the effects of a river basin development concentrate on energy production, those for natural gas fuel policy on the gas and petroleum sectors.

A Simple Econometric Model [23]

An econometric model views the nation (or region) by a series of flows shown in Figure 5. All flows are measured in dollars. As an example of an econometric model, we consider the simpler flow shown in Figure 6.

Personal consumption and industrial investment are the only sectors represented. The sum of these two demands is always met by the output of the nation—the national produce (gross and net being the same).

The behavioral assumptions are that each group of actors (consumers, industrialists) make demand decisions based only on past production levels.

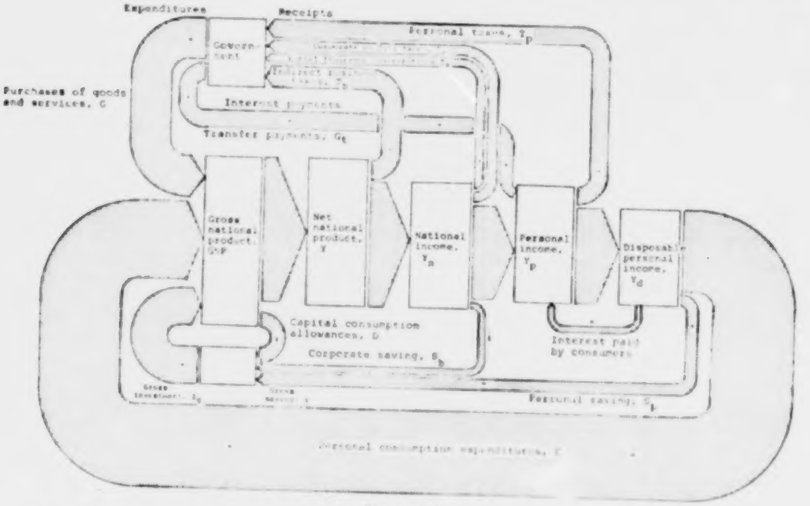

Expenditures
Receipts
Government
Personal taxes, T_p
Indirect business taxes, T_b
Purchases of goods and services, G
Interest payments
Transfer payments, G_t
Gross national product, GNP
Net national product, Y
National income, Y_n
Personal income, Y_p
Disposable personal income, Y_d
Capital consumption allowances, D
Corporate saving, S_b
Interest paid by consumers
Gross investment, I_g
Gross saving, S
Personal saving, S_p
Personal consumption expenditures, C

FIGURE 5

(Adapted from the chart "The Flow of Income and Expenditures in the United States 1964," published by the Twentieth Century Fund, Inc., 1965. The chart reproduced with minor changes with the kind permission of the publisher.

Consumers look at the last data available, the last period's production:

$$C_t = c_0 + c_1 Y_{t-1}$$

C_t is the consumption in period t.

Y_{t-1} is the national product in the previous period (one before t).

c_0 is the basic consumption (say, for "necessities") unaffected by past production levels.

c_1 is the *marginal propensity to consume:* that is, the increase in spending that occurs for a unit increase in disposable income, which in turn, in this case, is equal to the national product since there are no corporate retained earnings nor taxes. (This assumption on the way consumption changes with disposable income is only one of several that have been hypothecated, and is only an example.) c_1 is assumed to be greater than zero (never spend less when disposable income increases) and less than one (never increase spending more than the increase in disposable income).

Now we model the investment subsystem decisions:

$$I_t = i_0 + i_1 (Y_{t-1} - Y_{t-2}).$$

I_t is the demand or requirements by industry for goods, in period t, measured in terms of output.

i_0 is the basic requirement, unaffected by previous rates of production (e.g., for replacement of equipment).

$Y_{t-1} - Y_{t-2}$ is the change in product based on the latest information available (change from next-to-last to last period).

i_1 is the increase in investment (hence output) for a unit increase in the change in output. Notice that this behavior is related to changes, not levels. Businessmen are assumed to be optimistic and demand more when the product has been increasing rapidly, less when increasing slowly and actually have a negative demand (reduce output) when the product has been decreasing. (Many other assumptions are possible.)

The third relationship represents the market place: it says that supply equals demand:

$$Y_t = C_t + I_t.$$

Note that prices are not specifically represented, but are implied to have worked to bring about this equilibrium.

We now have the relationships needed to compute the way the product Y

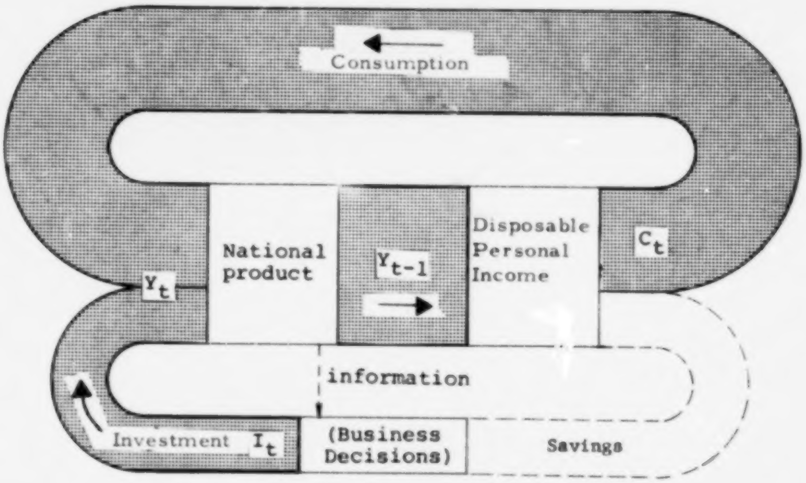


FIGURE 6—*Flow for Simple Econometric Model*

changes with time. If we assume continuity (very small, short periods) we can use calculus to compute the "time path" of Y. Or we can "simulate" by:

(a) assuming an initial value of Y_{t-1} and Y_{t-2} and then

(b) compute C_t from the first relationship

(c) compute I_t from the second

(d) compute Y_t from the sum of the amounts computed in (b) and (c) and

repeat b, c, d for as far into the future as we want to predict.

If we use calculus or run simulations with different parameters, we will see that there are four general shapes to the time path: (Figure 7):

(1) stable change to a new level,

(2) stable change to a new level with cycles or oscillations,

(3) unstable with cycles; product ever increasing,

(4) unstable without cycles.

The cycles are due to the fact that businessmen are assumed to act on changes and hence can overshoot the investment requirements.

All dynamic econometric models use the form of reasoning of this example. They just introduce many more subsystems and relationships. Typically they involve 15 to 150 equations.

Econometric models are so complex that the process of constructing them has become a discipline in itself. For most policy analysis, existing models are used. Construction of a model proceeds as described for simulation models, except that the designer consciously invokes economic principles in preparing the equations. He also estimated the parameters carefully, as we have said, using formal statistical methods; often developed by the econometrician.

Alternatives are introduced into the model by changing the appropriate parts and parameters. (If the models available do not explicitly contain the sectors affected by the proposed changes, a new model must be developed; a major effort. Since most existing economic models do not separate certain public functions, such as health, their use by government policy analysts has been limited.)

The econometric models are used in the same way as simulation models: the computer computes the predictions year by year.

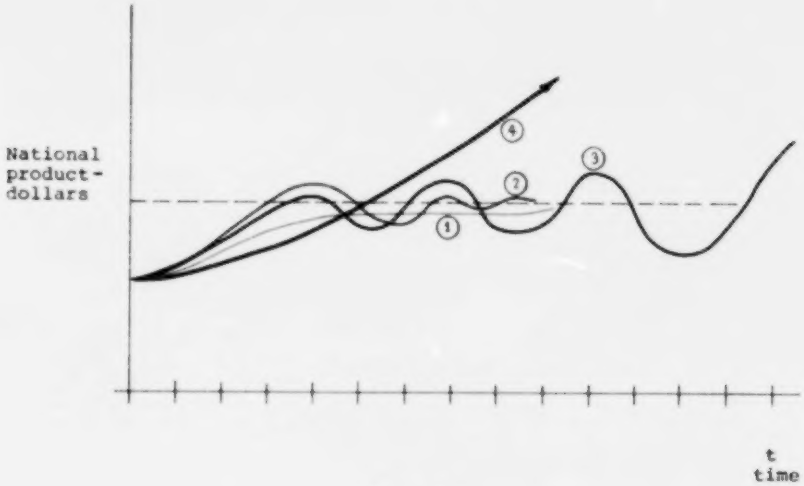


FIGURE 7

Input-Output Models

The basis of an input-output model is a table which shows the relationships, in dollar terms, between the inputs and outputs of each sector of the economy [36]. It is a highly-structured form of an econometric model (see Table 4). The entries in the table show for example, how much of the output of the steel industry is used by the automobile manufacturing industry and how much of its output is used in transportation services that are used by the steel and other industries. Also shown are the ultimate demand: inventory accumulation, exports, government and private purchases and capital formation. Inputs in form of payment from such sources into the economy are represented.

Input-output models are used to answer certain specific but important economic policy questions. For example, they can be used to make comparisons between nations. This could help underdeveloped countries determine the types of investments which would be most effective in stimulating growth. The input-output model can be used for studying the level or nature of the aggregate demand that would be required to achieve full employment (although full employment depends on other factors not always represented in the input-output model). These models are particularly useful in making a consistent short-term forecast; consistent in the sense that the inputs and outputs to each sector have the proper relationship to each other. France uses an input-output model as one tool for making forecasts of the requirements of each industrial sector. Input-output models can be used to study the effects of public programs such as public works or space programs on employment in industrial production.

To use input-output models for long-term forecasting requires some sophistication because they are basically static models. Some method of changing the factors within the model to reflect changes in technological developments, for example, are required if proper long-range forecasts are to be made.

Input-output models have been used to study regions of smaller size than the

Table 4. Hypothetical Input-Output Transactions Table *

Sector Purchasing (in Billions $)

Sector Producing	Outputs¹ → / Inputs² ↓	Processing Sector (1) A	(2) B	(3) C	(4) D	(5) E	(6) F	Final Demand (7) Gross inventory accumulation (+)	(8) Exports to foreign countries	(9) Government purchases	(10) Gross private capital formation	(11) Households	(12) Total Gross Output
Processing Sector	(1) Industry A	10	15	1	2	5	6	2	5	1	3	14	64
	(2) Industry B	5	4	7	1	3	8	1	6	3	4	17	59
	(3) Industry C	7	2	8	1	5	3	2	3	1	3	5	40
	(4) Industry D	11	1	2	8	6	4	0	0	1	2	4	39
	(5) Industry E	4	0	1	14	3	2	1	2	1	3	9	40
	(6) Industry F	2	6	7	6	2	6	2	4	2	1	8	46
Payments Sector	(7) Gross inventory depletion (−)	1	2	1	0	2	1	0	1	0	0	0	8
	(8) Imports	2	1	3	0	3	2	0	0	0	0	2	13
	(9) Payments to government	2	3	2	2	1	2	3	2	1	2	12	32
	(10) Depreciation allowances	1	2	1	0	1	0	0	0	0	0	0	5
	(11) Households	19	23	7	5	9	12	1	0	8	0	1	85
	(12) Total Gross Outlays	64	59	40	39	40	46	12	23	18	18	72	431

¹Sales to industries and sectors along the top of the table from the industry listed in each row at the left of the table.
²Purchases from industries and sectors at the left of the table by the industry listed at the top of each column.

* From Miernyk, Wm. H., *Input-Output Analysis*, Random House (5th printing 1967).

entire nation and even specific agencies, such as a hospital.

The preparation and use of input-output models are essentially the same as for any econometric model. Some simple alternatives can be analyzed by paper and pencil studies.

Mathematical Programming

In a situation to which mathematical programming applies, the following factors are assumed to be identifiable:

activities—processes or services which must be carried out to accomplish desired results,

resources—materials, consumables or processing capacity which is available in limited or *constrained* amounts,

outputs—the results of the activities; these are usually implicit in the levels of activities,

a *utility* measure—which indicates the overall value of producing the outputs,

coefficients—factors or parameters which measure how much of each resource is used for one unit of each activity, and

weights—which indicate the addition to the utility measure for a unit increase in each output.

In our earlier terminology, the levels chosen for the activities are the controllable variables. The maximum amounts of each resource, the parameters, and the weights are the uncontrolled variables. The parameters are usually determined by the technological characteristics of the processes, the weights by economic factors and the resource constraints by previous capital expenditures or resource limitations. The relationship between outputs (and weights) and utility is the objective function [22].

As an example of mathematical programming from the public sector, the activities might be the application of controls on pollution. The more controls that are introduced the higher the level of the activity. The resources used are funds, materials and space constraints which limit the extent to which controls can be introduced. The outputs are the reductions in pollutants. The utility is the benefits of reduced pollution: better health, etc. Utility might be estimated by assigning weights to reductions in pollutants which gives an overall estimate of the value of the reduction. For example, it might be more important to reduce sulfur dioxide than particulates. The utility function then might contain weights such that sulfur dioxide reductions were weighted by a higher number than particulates.

For every unit of control introduced a certain amount of each of the resources is used. The coefficients indicate the extent of this usage. If the utility and resource relationships are all linear (that is, there are no economies of scale, no diminishing returns or other such effects) then the situation can be represented by a *linear* programming model. Optimizing solutions for such linear cases can almost always be found. Where one or more of the relationships are not linear, then we have a *non-linear* mathematical programming problem. Optimizing solutions can not always be found for these. However, stating a problem in the form of a non-linear mathematical programming model may be useful in suggesting the nature of the optimum or indeed in finding one by some special means.

Mathematical programming models are highly structured, unlike simulation, and are generally applicable only in well defined cases where resources are allocated to pre-defined programs or activities.

With linear situations, computer programs already exist which find the optimum. The analyst merely has to choose the proper computer software, introduce the data which represents his model and submit the program to the computer. There may be the usual difficulties in getting the data in the right form. But the use of linear programming is generally simpler than for other models.

For nonlinear situations a few specialized optimizing programs exist. If the analyst can use these he can usually expedite the development of the optimizing process. Solutions to nonlinear problems, however, use large amounts of computer time and, therefore, are not always justified.

A good introduction to linear programming is reference [22].

Decision Trees

Decision trees are a simple form of simulation which is structured by aggregating future activities into major decisions and their consequences [6]. The model represents only three basic variables: (1) the probability of each of the possible consequences of decisions, (2) the cost and (3) the benefit, usually measured in dollars of each possible consequence. The purpose of a "tree" is to estimate the expected or probable value of taking each of several alternatives at intermediate decision points. A tree can incorporate estimates of the immediate and future costs that will arise because of the decision and of the consequences of future decisions.

A decision tree may be looked at as an overall model of the possible consequences of a decision. Other more detailed models would be used to obtain the estimates of the probabilities, the costs and the possible benefits from each of the alternatives. Figure 8 shows a typical decision tree, describing a simple and more detailed model of the possible alternatives and consequences of a pollution control or education program.

The main mathematical process used in the decision is Bayesian statistics. Initially subjective estimates are made of the probabilities of various consequences. Then the Bayesian statistical method is used to get more refined estimates of the probability of the combinations or sequences of activities. This gives the net estimate of benefit by computing the probable total benefit minus the probable total cost of each alternative.

One of the main uses of this analysis is to help determine the value of *additional information* about a situation. In order to do this it is necessary to have some estimate of the probable outcome an information gathering effort, e.g., a major survey. This combined with the basic decision tree information, allows an estimate of the probable change in the overall outcomes, if the information were to be collected. Since it is usually easy to estimate the cost of the information collection, it is possible to weigh this against the probable change in the benefits to determine if the information collection effort is worthwhile.

The main assumptions in use of decision trees is that the future can be summarized by a sequence of activities followed by decisions and that the alternative consequences of each decision can be identified. It also assumes that the required data can be estimated.

Construction of the tree is extremely straightforward. It involves simply drawing the tree on the basis of information available about the alternative sequences of events that are likely to take place in the future. The sequences of events can be derived from either the results of another modeling effort or on subjective estimates of the decision maker and staff.

In some cases data about the probability and costs of each activity is available from other studies. Often in a decision tree analysis, the estimates are made by the subjective input of the staff, experts, and the decision maker. Various techniques have been developed to elicit from decision-makers estimates of the probability of alternative events occurring in such a way as to obtain fair and unbiased estimates.

The computations for a tree can usually be done by pencil and paper although computer programs might help in a very complicated tree.

A decision tree analysis might be a good model to use both at the beginning and at the end of a more exhaustive analysis. At the beginning it can help pinpoint the decisions that need to be made, their probable consequences and to determine whether the overall analysis effort has any value. At the end it is a useful device for summarizing the results of detailed studies to get again a picture of the overall cost-benefits from the decision.

III. HOW MODELS ARE CREATED AND TESTED

In order to clarify the nature of models and of the process of modeling, this section presents a review of the steps which an analyst may go through to create and test a model, and then to use it to help the decision maker. All of the steps which can occur will be mentioned, but not all of them are required in any one case. In some cases the tested model exists and its use can start almost immediately. (For application of the process to specific problems refer to the literature cited at the end of this chapter and to the use of models described in the other chapters.)

Recognizing the Need for a Model

We cannot give a rule for recognizing the need for a model in any given situation. Up to this point in our discussion we have asked the reader to assume that a special and important class of decision-making aids, termed models, has been shown to be of value in making decisions of substance. We are the first to recognize that this assumption is not always true. There are many specific problems for which models as presented in this volume do not provide useful inputs. We do feel, however, that decision makers and their staffs must be in a position to determine if and how a decision problem can be resolved more effectively by the analytical power of a model. As part of the attack on any such problem, there should be a conscious choice to investigate or to not investigate the suitability of a model. This chapter, and the descriptions in the succeeding chapters of how models have been of value, is directed towards improving the quality of this conscious choice. Decision makers and their staffs must rely on their previous experiences, and on professional contacts, reading and training to determine those situations for which models and related methodology are applicable.

Suppose that the decision maker becomes convinced that a model could be helpful. What happens next?

Analysis often begins as a result of a specific management directive. When the problem-formulation phase is initiated through management's actions, the assignment may be vague or quite specific; if it is specific it may be right or wrong. The analyst might be told to "see if we can reduce the cost of operating our vehicles" or he might be specifically told that "we need another vehicle, prepare a justification for it." The second directive might actually be wrong, since a detailed study could show that better scheduling and routing rules would increase production as much as adding a vehicle and would do so more cheaply. When the analyst receives very specific directives, he is under some organizational pressure to carry them out. However, he should be

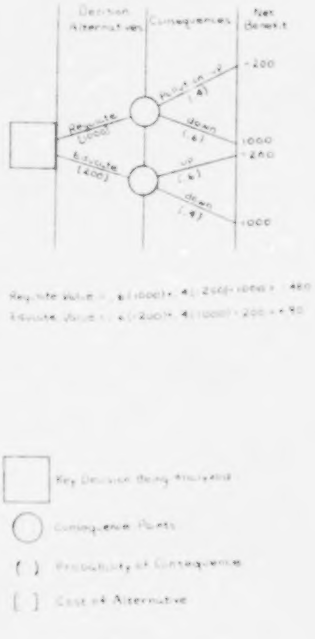


FIGURE 8—Decision Tree Examples

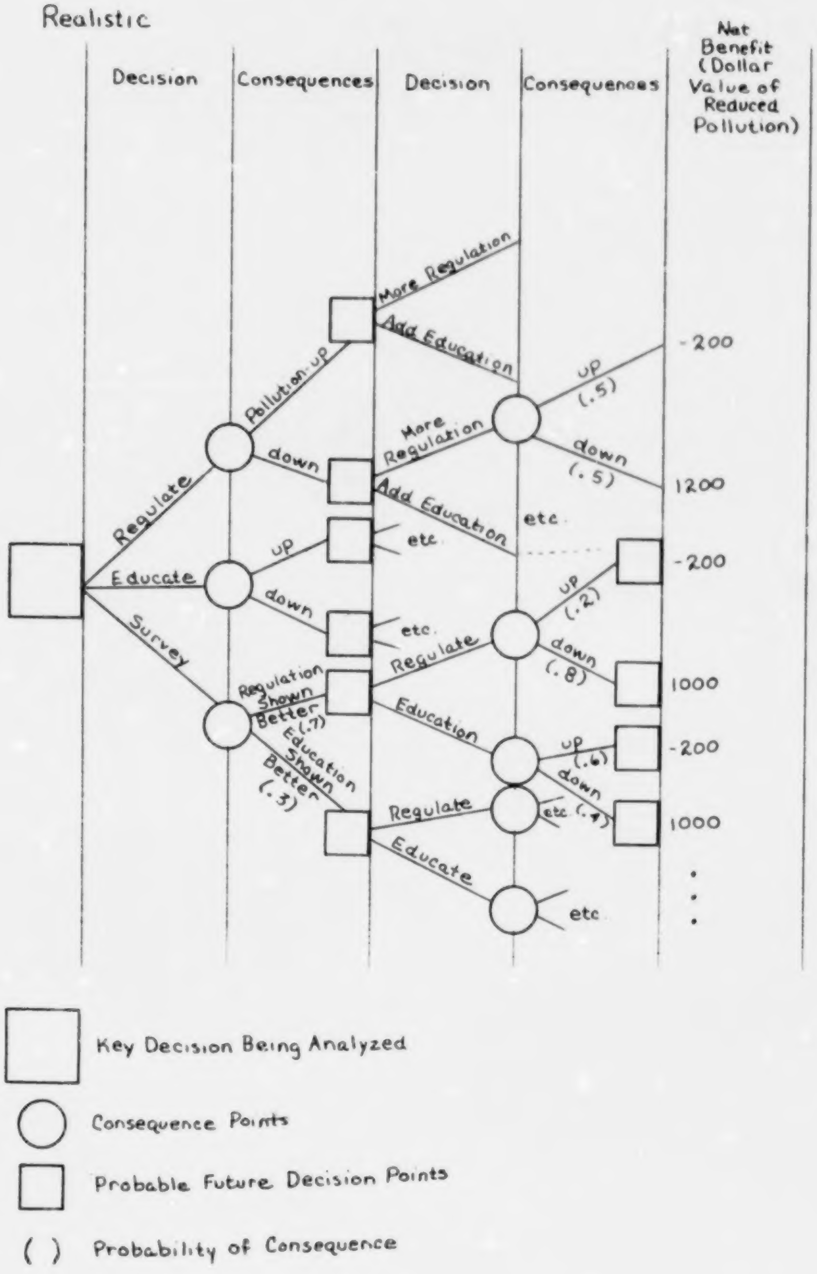

Realistic
Decision
Consequences
Decision
Consequences
Net Benefit (Dollar Value of Reduced Pollution)
More Regulation
Add Education
Pollution up
up (.5)
-200
More Regulation
down (.5)
1200
down
Add Education
etc.
Regulate
up
etc.
-200
Educate
down
etc.
up (.2)
Survey
Regulate
down (.8)
1000
Regulation Shown Better (.7)
Education
up (.6)
-200
Education Shown Better (.3)
down (.4)
Regulate
etc.
1000
Educate
etc.
Key Decision Being Analyzed
Consequence Points
Probable Future Decision Points
( ) Probability of Consequence

FIGURE 8—Continued

aware that doing this may result in a suboptimal solution. If preliminary analysis indicates that this is the case, the analyst should be sure the decision maker is willing to forego the chance of a better solution.

It is important to keep in mind that the first time a model is used within an organization is quite different from subsequent uses. The first time usually involves a large investment; subsequent uses are less costly. The first time there may be unresolved questions about the validity of the model; its output should be taken only as guides. Each use of the model should improve its validity and acceptability so that eventually its output may be taken as a decision to be modified occasionally. Obviously the more complex the situation, the longer before the model becomes fully acceptable (and in many policy situations this may never occur).

Let us look at the steps involved in the first-time use of a model. Fig. 9 indicates how the process of selecting a method of analysis might proceed.

Problem Definition

The process starts by defining the problem: a complex and creative activity. It involves determining the boundaries, at least tentatively, of areas of concern. It requires the definition of the following:

- the decision; specified by identifying those factors or variables which the decision maker can control: the controllable variables.
- measure(s) of performance by which the value of the decision is to be judged.
- an objective function to relate certain output variables to the measure of performance; it tells how to estimate the relative value of the outcome of any decision.

Choice of Analysis Method

After the problem is defined, the analyst asks, "Does any standard model fit the problem?" If one of the standard models does fit, the appropriate solution can be found very efficiently since procedures for solving most standard mod-

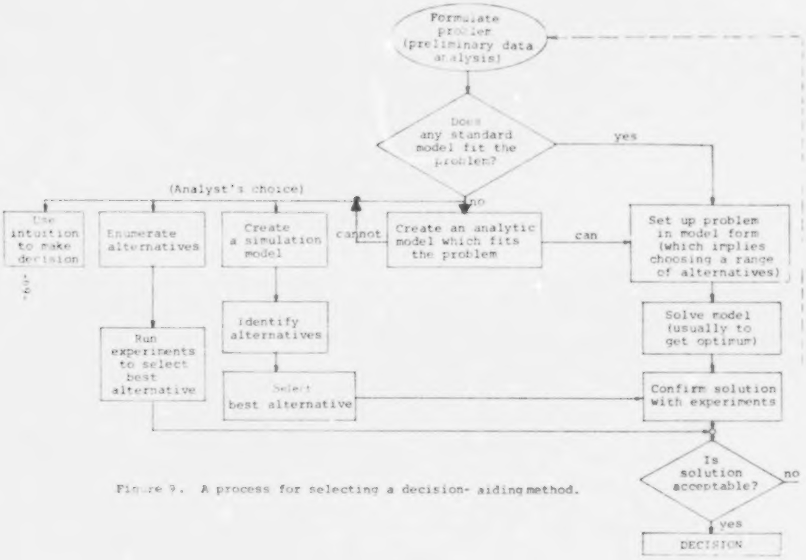

Formulate problem (preliminary data analysis)
Does any standard model fit the problem?
yes
no
(Analyst's choice)
Use intuition to make decision
Enumerate alternatives
Create a simulation model
cannot
Create an analytic model which fits the problem
can
Set up problem in model form (which implies choosing a range of alternatives)
Identify alternatives
Run experiments to select best alternative
Solve model (usually to get optimum)
Select best alternative
Confirm solution with experiments
Is solution acceptable?
no
yes
DECISION
Figure 9. A process for selecting a decision- aiding method.

FIGURE 9—*A process for selecting a decision-aiding method.*

els are known. In this case, the right-hand sequence of steps in Fig. 9 is followed:

(1) The problem is set up in the standard form.

(2) Data are gathered to provide the inputs for the model. Specific parameter values are estimated, sometimes using statistical techniques.

(3) Any necessary adjustments are made in the model to account for the particular situation.

This process is even easier if there is a case where the model has been applied to a similar situation. This reduces the work in Steps 1 and 3. One purpose of this book is to make the reader aware of standard and special, but well-developed models.

From the decision maker point of view, it is a critical step to determine whether there is a standard model or not. If there is, the cost of the modeling effort is relatively low and it is worth assigning analytic staff to carry the modeling process forward (at least until good cost estimates can be made). If not, the development of a model should, perhaps, be explored with others who have similar problems before making any expensive commitment.

The model is solved by deducing which is the best of the range of possible decision variables, using the optimizing or search processes associated with the standard model. In a complete operations research study, the solution derived through the model would be confirmed by field tests, where possible, and this solution, if it is acceptable to management, is implemented.

If none of the standard models fit the situation in question (which is a common occurrence), then the analyst has the following alternatives.

He can try to create a suitable mathematical optimizing model. This may involve applied mathematics, and in many cases, behavioral research. Such analysis is a high-risk alternative; a suitable model may not be created within the time frame available for decision making. However, if a model can be created, then the steps on the right side of Fig. 9 apply. If not, one of the other alternatives must be chosen.

When no standard model fits, the analyst may choose to use simulation. A special simulation model, which is a predictor, must be designed. The first step here would be to determine if there is a model already developed that represents a similar situation. If so, adapting it may be the most effective approach. Simulation models are, in general, fairly costly and require a major commitment. Computer simulation languages have been developed which aid in reducing the cost. To use the simulation, the alternative solutions are identified, and the simulation model is used to test for the best alternative. When found, the alternative may be confirmed by field tests.

If the analyst decides not to use a simulation model, two approaches remain, direct experimental and intuition. In the former, alternative courses of action are identified and the best is selected by direct experimentation in the real world. Finally, if none of these approaches are used, the analyst and the decision maker can resort to intuition.

The Basic Steps in Model Design

Figure 10 is a flow chart of the major steps that could be taken in the design of a model (or a major adaptation of an existing model) [4]. We describe these steps in general terms in that an understanding of the process by the decision maker will facilitate greatly his ability to interact with model developers.

Notice that at a number of points in the process may return to a previous step. This illustrates one of the important characteristics of successful modeling: the earlier steps are always subject to revision as additional information is obtained. For example, a review of data often causes some reformulation of the problem; especially, as is often the case, some data are unobtainable. Actual runs of the model almost always suggest new alternatives, and therefore, a return to alternative definition.

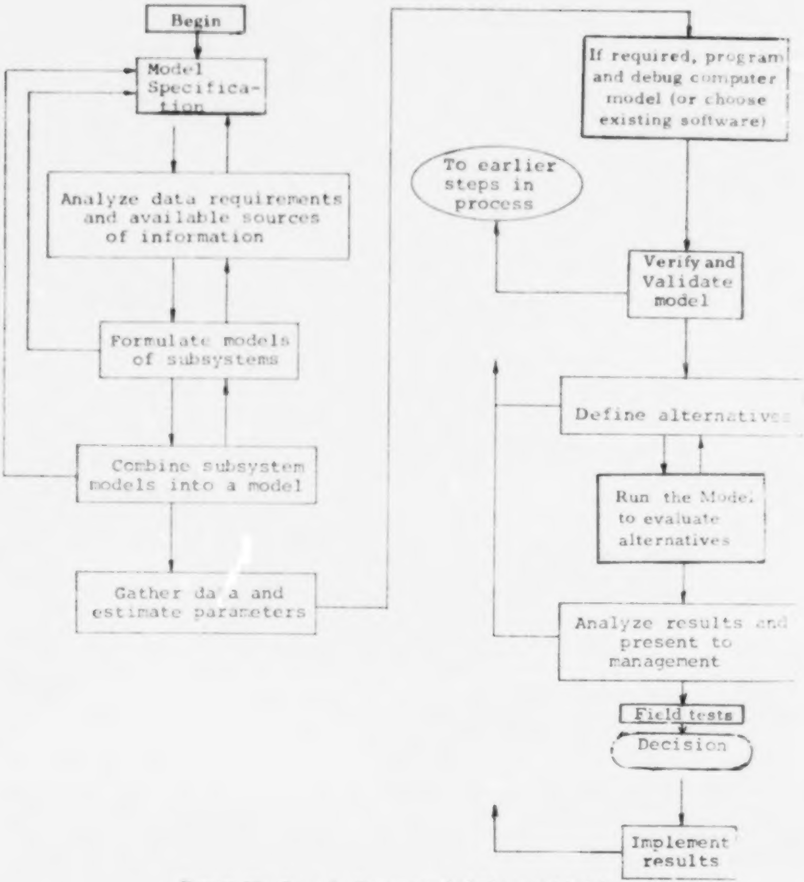

Begin
Model Specification
Analyze data requirements and available sources of information
Formulate models of subsystems
Combine subsystem models into a model
Gather data and estimate parameters
If required, program and debug computer model (or choose existing software)
To earlier steps in process
Verify and Validate model
Define alternatives
Run the Model to evaluate alternatives
Analyze results and present to management
Field tests
Decision
Implement results

FIGURE 10—*Steps In Designing And Using A Model*

Model building is as much an art as it is a science. One of the difficulties in using a flow diagram to represent the model building process is that the process looks more scientific than it really is. The flowchart in Fig. 10 is only a guide and should not be interpreted as a method that automatically produces the creative leaps required to translate a complicated, real world system into a more compact and manipulatable model.

Model Specification

Specification is the most critical step in developing a model of a system. An analyst's success in accomplishing other research objectives depends on correct formulation.

The first step is to specify the objectives to be achieved through the analysis. These objectives will generally be stated by management after they recognize that some aspect of the system is not functioning to their satisfaction. The problem, as management sees it, is likely to be defined vaguely (from the analyst's point of view), and the objectives stated in qualitative terms. It is the analyst's job to translate these qualitative objectives into operational terms that can be related to output variables.

For an optimizing model this will include the creation of an objective function. Usually the analyst will create this function by assigning weights to each output parameter or result according to its importance to management.

Along with specifying the output variables and an objective function, the analyst tries to identify relevant variables in the system; that is, variables that effect the outputs. He then separates the relevant variables into two classes: those having values that can be controlled by the decision maker and those which cannot. The analyst's goal will be to choose the levels of the controllable variables with the aid of the model, in a manner that optimizes as nearly as possible the stated objective function. The distinction between controllable and uncontrollable variables is not always clear. Depending on the scope of the study, variables are sometimes treated as uncontrollable, although the decision maker has (or has potentially) some control over them. The behavior of people in generating trash is uncontrollable in the short run and is so considered in sanitation truck scheduling studies. But this behavior is controllable by educational programs or incentives.

Constraints are uncontrollable constants, e.g. upper limits on resources available. (We have treated problem definition and model specification as if there were a clear-cut sequence in which the analysis occurs: the steps are really intertwined.)

Knowledge of existing models and of theories is an important source of guidance for problem formulation. (Philosophically, is it even possible to describe any problem situation precisely without some model in mind?) For example, most analysts who are faced with an allocation problem use the ideas of the standard model called linear programming to initially organize the variable relationships, constraints, and system objectives in their minds. This thought process influences the way the problem will be formulated, even when the actual situation does not satisfy linear-programming assumptions.

When the analyst has set the specification, he begins the task of model construction. Model construction usually begins by the identification of subsystems and variables in the subsystems. General rules cannot be given for segmenting a large problem into subsystems; this is almost entirely dependent upon the structure of the system being analyzed. Subsystems should be as independent of each other as is possible. The size of a subsystem necessary to meet this requirement depends on the amount of interaction between the variables in the total system.

The extent to which observation and previous experience assist in the understanding of the subsystems varies. An analyst with experience in office procedures will have little trouble formulating subsystem models about operations of a specific office. However, even an experienced welfare analyst could have trouble formulating relationships that predict the effect of a change in welfare qualification rules. One reason for this is that such predictions require an understanding of recipient behavior, and less is known about representing behavioral processes than about representing information and office processes.

Perhaps the most important point about model building is that the analyst should be intimate with the situation and its theoretical basis, either from previous experience or by intensive study. To successfully model a pollution situation, say, it is necessary to understand chemistry, thermodynamics, fluid flows, supply-demand-price theories, etc. To simulate the behavior in a particular welfare situation, it is necessary to understand what is known about relevant psychological and sociological theories of behavior, as well as to know how the social services are provided.

Analysis of Data Requirements and Availability of Sources

In the model specification step all parameters and variables that effect the measures of performance were identified. Now the analyst determines what data are needed.

For modeling, data are needed to:

- estimate values of constants and parameters,
- provide starting values for all variables,
- provide data to which outputs can be compared for validation: historical values of key variables.

To ensure that this data will be available sources of data are located and their adequacy is evaluated.

Creation of the Model

Even though the model is of a standard type or one that has been developed elsewhere, the analyst may still have to adapt it to the particular situation. This may be a major effort in manpower and time, although conceptually less difficult than designing a model *de novo*. It is an effort that can consume several man-months for a simple case, like the vehicle replacement problem, or dozens of man-years, in policy or major allocation cases, like the natural gas regulation or pollution problems.

As just discussed, various types of data are needed to test, operate and validate the model. A critical step, and often the most costly step, is the mundane process of getting the data. The model serves as a guide to what data are required: this is one of the advantages of having a model. The activities required are:

- Negotiate with sources of data to obtain it: in public applications, the data are often in the files of many different agencies. Privacy or confidentiality considerations may make data acquisition difficult.
- Obtain the data from the sources: this presents two kinds of problems: (1) the economics of extracting data and (2) the political problem of gaining the cooperation of an agency to use its data. In regard to the first problem, if the data are not already in machine-readable form (cards or tape) it may not be worth the effort to use a model that requires the data. In some cases estimates of parameters and initial values can be made with small samples of non-machine readable data.
- Edit the data and performing statistical operations to estimate parameters or to forecast future values of variables. (This can involve major computational and data processing efforts).

Computation: Programming and Debugging

Since a model consists of a set of variables and their logical and mathematical relationships, a computer may be appropriate for analyzing alternative solutions. If this is not the case, then hand computational procedures, tables and other aids may be all that is required. Generally, programs for standard models are already available in pretested software, i. e. available computer programs. A model adapted from someone else's work probably exists in programmed form, but may need adaptation. If neither of these are the case, then the analyst must prepare the corresponding computer program. Sometimes an aid to the preparation of new models is available in the form of an analyst-oriented, computer language. There are several such languages which are quite helpful in reducing the time and cost required to formulate, prepare, test and debug models, analysts should be aware of these.

Model Verification and Validation

Further steps in model development include verification and validation. The former is concerned with determining whether the statement of the problem, e.g. the computer program, represents the desired model under all anticipated conditions. This might include using the model with input data which correspond to known output measures. The validation step seeks to establish how close the model mirrors reality. In a few cases the analyst can compare the model's output measures against historical results or the results of a field experiment.

Model Use

At this point we have a tested model program and data. If the model is an optimizing form, then it is now ready for use: the data are input and the program run to produce the outputs. Most such programs not only produce the optimum values of the decisions variables, but also *sensitivity* information. This might include an indication of the costs of deviating from the optimum, and the benefits that might be derived by relaxing constraints, that is by changing factors originally assumed to be "uncontrollable."

If the model is only a predictive model, then one more step is needed: it is necessary to decide how it will be used to search for the best decision. The search process may be:

- try several alternatives, or
- design a simulated "experiment" to try alternatives in a controlled way, which insures trying a wide range of them, or
- try a number of alternatives at random, or
- use a formal search process in which the next alternative to try is computed so as to lead to a higher utility.

Analysis and Presentation of Results

The results of a model are seldom completely consistent with preconceptions of the outcome. The study should provide new insights into the interrelationships between the components of the system. There have been cases where insights into new alternatives, derived while designing and preparing to use the model, were the most valuable results of the process. Therefore, it is critical that the presentation of the results to management are in a form that maximizes this insight. *Ideally, the decision maker will be involved throughout the process and would have both contributed to and gained from these insights.*

Since it is unlikely that the analyst has been able to make a perfect translation of the objectives into a model, the decision maker is wise to generate his own alternatives for analysis. The analyst can improve the objective function by interacting directly with management concerning the relative value of the alternative solutions.

Implementation

Any analysis should be looked at from three points of view. First, it should provide specific operational recommendations which can be implemented. Second, it should provide decision makers with additional knowledge and understanding of their organization and environment. This knowledge may uncover important, but currently unrecognized, problems that require further analysis. Third, the results of any study should be used to improve the model and the experience of the analyst.

Documentation

Documentation is a vital part of the process of developing and implementing a model. Documentation serves as the basis for communication between the various people who must be involved in the modeling effort if it is to be successful. Such communication paths must be created between the decision maker and the analyst, and between the analyst and other personnel affected by the study.

In some cases there is also communication from the analyst to the profession in general (i.e., through journal articles, professional meetings, and so forth) so that others may improve practices as a result of individual experience.

One thing that must be documented is the definition of the problem. This includes:

- a description of the organization in which the problem exists,
- a statement of the decision that is to be made and the range of alternatives possible (to the extent that they are presently recognized),
- a list of the constraints and guidelines considered to be outside the problem area, which form the set of assumptions under which the analysis is being done,
- the model specification,
- a complete statement of the basic nature of the processes in the system.

The documentation should also include the results of the analyses, recommendations and some statement on how the model should be revalidated and updated based on the results of implementing the decision maker's choice of solution.

IV. PROBLEMS IN USING MODELS

The purpose of this section is to bring together a discussion of the problems that can arise in attempting to use a model.

Defining Measures of Effectiveness

It would seem that a decision maker should know what he is striving for and how to measure progress towards his goals. It turns out, however, that, when an effort is made to define specific measures of progress, difficulties arise. Measures of progress are also called "measures of effectiveness," "measures of performance," "criteria," or "indicators." An objective function is a relationship between certain observable and measurable phenomena to compute a single overall measure of effectiveness.

The kinds of difficulties that arise are the following:

- Agreement cannot be reached on what the measures of effectiveness should be.
- Certain objectives cannot be quantified.
- Quantifying measures can be stated, but there is no feasible way of making the measurements to determine their value at any given time.
- There is agreement on a set of several measures but no way of combining them to form a single overall measure of effectiveness.

There are methods for handling multiple criteria [9, 10, 32]. It is a difficult problem, however, and it arises frequently in government where most programs do have more than one goal.

To the extent that the problems just mentioned are a reflection of the fact that the participant decision makers truly have different values, the solution must lie outside of the modeling process. To the extent that the problem is one of communication and definition, it is usually solved by meetings among the relevant people in a situation conducive to good communication and exchange. To the extent that the problems are economic or technical, such as that there is no feasible way of making the measurements implied by the specified criteria, then either surrogate indicators must be sought or else the modeling effort given up. The use of surrogate measures implies that the decision maker is taking on faith a relationship between the desired measure and the surrogate indicator.

Aggregation

In most cases the model, which is an abstraction of the real world, does not represent every person, event or activity that could be relevant; rather it represents aggregations or groupings of these. A critical problem in the design of a model is to decide to what extent these groupings should occur. If the aggregation is too great, then certain detailed phenomena which can affect the predictions and decisions may be lost. If the aggregation is too fine, however, the model becomes unwieldy and unusable.

Data Acquisition

This is largely an economic problem. A few models may call for data which are inherently unobservable (this is particularly true for certain behavioral phenomena). Most models have been designed so that the data required to establish parameters, coefficients, initial values and so on can, in principle, be observed and measured. The problem is that the cost of making these measurements may be so high so as to make the modeling effort infeasible.

In some cases a modeling effort has led to the recognition of the need for an information system. The modeling itself then must await the implementation of that system.

There are other problems related to

the use of data, having to do with the invasion of privacy or confidentiality. Some data may not be available because divulging them would either compromise information about individuals or about the organization and how it operates. In the long run, organizational data about most public organizations can be made available. However, the problem of individual privacy is still a major one. Often models require only statistical or aggregate data so that individuality can be eliminated at an early stage of processing and so avoid the privacy problem. [34]

Verification and Validation

The computer program which represents the model must be thoroughly tested to insure that it does represent the desired model. There are cases known where a model has been used for several years before certain errors have been found which indicate that the model has not been giving completely correct results.

A more serious problem is in determining the validity of the model. [4]:

> "Suppose, an analyst feels his model is an accurate representation of a system and he presents it to management for their use in decision making. The latter then asks, 'How do we know it is valid? How do we know its predictions will come true?' These questions must be answered at some point in every study.
>
> How can we be sure that the predictions made by any model will be correct or at least will be better than the prediction made by some other method (e.g., by judgment)? Ultimately this becomes a question of the credibility of the model. Therefore, we must examine what evidence is required by a manager before he will utilize a model as an aid in his decision making. Mathematical models have an advantage in this regard over simulation models. If, for example, a situation can be shown to fit the assumptions of a linear programming model, then the manager has the support of evidence that shows that linear programming models have proven effective in other (albeit different) situations. The only possible evidence of validity for a simulation model that has been developed specifically for a situation is that the model has made satisfactory predictions in the past. If this is the first time the model is being used, such evidence is not available. This difficulty is most severe with a simulation model of a nonexistent system, for the analyst cannot even test it using historical data."

Recognizing that it is never possible to completely validate a decision-aiding model, since there are never real data about the alternatives not implemented, we suggest that the validity of the model should be determined by the application of the following steps:

(1) The analyst assures himself that the model performs the way he intends it to, using test data, and if available, real historical data.

(2) Reasonableness is checked by:

  (a) showing that key subsystem models predict their part of the world well (using historical data).

  (b) showing that the parameters have reasonable values.

  (c) having people who are knowledgeable about the situation (preferably including the decision maker) review the model in detail and agree to its structure and parameters.

(3) The decision maker has an opportunity to explore the use of the model to become familiar with its predictions and to examine the interactions it implies. At this point the analyst and decision maker may be able to agree as to what is a close enough fit between model output and actual data.

(4) The model is used to aid decisions. Careful records are kept of its predictions and of actual results. (This may involve a time span of years, so that the evaluation procedure has to be set up carefully.)

Cost and Personnel Requirements

The use of a model can be very expensive, particularly the first time. A typical modeling effort in a non-trivial situation can easily consume two or three staff years, plus thousands of dollars of computing time and require one or more years of calendar time. This is especially true if the model is designed for a particular application, but can be equally true when the model is of a standard type or is one built by others.

Often the personnel required for these efforts is time of staff already on board and, although it displaces other analytic projects (a significant opportunity cost), it does not show up as a specific budget item. The computing time required to test and run a model, however, usually is a very specific item and therefore gains more attention that it deserves. Models can consume thousands or tens of thousands of dollars of computer time if the work is not carefully controlled.

Although a modeling effort may be easily justified by the savings and increases in effectiveness gained through better decisions, it is a difficult investment to justify in advance. Thus the decision maker must allocate funds for the initial model development effort somewhat on the basis of his judgment and faith in the lead analyst.

Setting Up the Model

One of the more difficult modeling processes is to translate the real world into the model. This problem arises when the decision maker wants to define an alternative which does not exist and wishes to analyze this alternative via the model. The decision maker defines the alternative in terms of such factors as policies, techniques to be used, and personnel assignments. These must be translated into specific model parameters. This step is a creative process on the part of the analyst which must be understood fully by the decision maker when he interprets the results of the model.

Interpreting Results

The outputs of a model must be interpreted into plans for action. Most decision makers wisely recognize the model outputs as a guide and not as gospel. Nevertheless, having invested a great deal of effort in a model and having observed its very precise (but not necessarily valid) outputs, the model results may weigh heavily in the decision making. Therefore, it is important to understand the effect of the choice of the model on the decision process. Since every model is an abstraction and emphasizes certain aspects of the real world and not others, different models will have different influences. The decision maker should consider this when interpreting the results of the model. The decision maker must fully understand the modeling process so that he can either change the model to reflect his views or interpret the results appropriately.

References

1. Richmond, Samuel B., *Operations Research In Management Decisions*, New York, Ronald, 1968, pp. 251–264.
2. Kohn, Robert E., "Application of Linear Programming to a Controversy on Air Pollution" *Management Science, 17*, No. 10, (June 1971), pp. 609–622.
3. Sundstrom, D. E., Review of [12] *IEEE Trans. on Sys. Sci. and Cybernetics*, Vol. SMC-3, No. 1., (Jan. 73), p. 112.
4. Emshoff, J. R. and R. L. Sisson, *Design and Use of Computer Simulation Models*, New York, Macmillan, 1970, pp. 5–11.
5. Ackoff, R. L. and B. H. P. Rivett, *A Manager's Guide to Operations Research*, New York, Wiley, 1964, or for more complete and advanced discussion, Ackoff R. L., and M. W. Sasieni, *Fundamentals of Operations Research*, New York, Wiley, 1968.
6. Hare, Van Court, *Systems Analysis: A Diagnostic Approach*, Harcourt, Brace World, 1967.
7. Churchman, C., Managerial Acceptance of Scientific Recommendations" *California Management Review*, 7, No. 1, (Fall 1964), pp. 31–38.
8. Ackoff, R. L., *Scientific Method Optimizing Applied Research Decisions*, New York, Wiley, 1962, Chapter 3.
9. Charnes, A. and Cooper, W. W., *Management Models and Industrial Applica-*

tions of Linear Programming, Vols. I and II, Wiley, New York, 1961.

10. Lee, S. M. and E. R. Clayton, "A Goal Programming Model For Academic Resource Allocation" *Management Science 18*, No. 8, (April 1972), pp. B395–408.
11. Wilde, D. and C. Beightler, *Foundations of Optimization*, New Jersey, Prentice-Hall, 1967.
12. Steele, J. L., *The Use of Econometric Models by Federal Regulatory Agencies*, Heath, Lexington, 1971, pp. 31–33.
13. Forrester, J. *Urban Dynamics*, Cambridge, Mass., MIT Press, 1969.
14. Thompson, S. P., *Calculus Made Easy*, Macmillan, London, 1946.
15. Moroney, M. J., *Facts from Figures*, England, Penguin (A234).
16. Yeomans, A. E., *Introducing Statistics* (Vol. I), *Applied Statistics* (Vol. II), Penguin (X5, X6) 1968.
17. Levin, R. I. and C. A. Kirkpatrick, *Quantitative Approaches to Management*, New York, McGraw-Hill, 1971 (revised ed.), Chap. 14.
18. Sisson R., et al., *Management Science Selections*, Los Angeles, Data Processing Digest, 1965, p. 39.
19. Mirasol, N., "A Queuing Approach to Logistics Systems," *Operations Research*, Vol. 12, p. 723.
20. Cox, D. R. and W. L. Smith, *Queues*, Methuen Mono., New York, Wiley, 1961.
21. Brown, R. G., *Statistical Forecasting for Inventory Control*, New York, McGraw-Hill, 1959.
22. Gass, S., *An Illustrated Guide to Linear Programming*, New York, McGraw-Hill, 1970.
23. Kogiku, K. C., *An Introduction to Macroeconomic Models*, New York, McGraw-Hill, 1968.
24. Samuelson, P. A., ed., "Mathematics in Economics," *Review of Economics and Statistics*, 36, (Nov. 1954) pp. 359-386.
25. Samuelson, P. A. *Economics, An Introductory Analysis*, New York, McGraw-Hill, 1967.
26. Wein, H. H., *Natural Gas Supply and Demand*, Washington, D.C., Fed. Power Comm. Docket No. AR61–1, (1963) p. 6.
27. Duesenberry, J. S., et al., *The Brookings Quarterly Econometric Model of the U.S.*, Rand McNally, 1965.
28. Liebenberg, M., et al., *A Quarterly Econometric Model of the U.S.*, Survey of Current Business, 1966, pp. 13–39.
29. Amstutz, A. E., *Computer Simulation of Competitive Market Response*, Cambridge, MIT Press, 1967.
30. Meadows, D., et al., *Limits to Growth*, New York, Universe, 1972.
31. IBM Corp., *General Purpose System Simulator 360*, User's Manual, H20-D326-0, 1967.
32. Courtney, J. F., et al., "A Goal Programming Approach to Urban-Suburban Location Preference" *Management Science*, Vol. *18*, No. 6 (Feb. 72), pp. B258–268.
33. Sang, M. L. and F. R. Clayton, "A Goal Programming Model for Academic Resource Allocation" *Management Science*, Vol. *18*, No. 8 (April 72) pp. B395–408.
34. Anderson, R. E., and E. Fagerlund, "Privacy and the Computer: An Annotated Bibliography," *Computer Reviews*, Vol. *13*, No. 11 (Nov. 1972), pp. 551–559.
35. Hermann, C., "Validation Problems in Games and Simulations" *Behavioral Science*, Vol. *12*, (May 1967), pp. 216–230.
36. Miernyk, W. H., *The Elements of Input-Output Analysis*, New York, Random House, 1967.

Chapter 2

Models and Policy Making

By

Peter W. House
Gene R. Tyndall

I. INTRODUCTION 41
  Purpose 41
  Scope 41
II. POLICY MAKING DEFINED 42
  Levels and Scope of Policy Making 43
  Scope of Problems 45
III. POLICY-ORIENTED MODELS 47
IV. NEEDS OF POLICY MAKERS 49
  Interviews and Questionnaires 49
  Questions About the Policy Maker 49
  Questions on the Policy-Making Process 50
  Questions on the Use of Models in Policy Making 51
V. OPPORTUNITIES FOR MODELERS AND POLICYMAKERS 51
  Challenges to Modelers 52
  Challenges to Policy Makers 55
VI. CONCLUSION 55
REFERENCES 56
APPENDIX—AN EXAMPLE: THE STRATEGIC ENVIRONMENTAL ASSESSMENT SYSTEM (SEAS) 56
  Scope of SEAS 56
  Residuals 58
  Effects 58
  Reactions 58
  Evaluators 58
  Expected Uses of SEAS 59

Preceding page blank

Models and Policy Making

I. INTRODUCTION

Throughout the chapters of this book, especially within the Primer section, there are implied assertions that models are "natural" aids for decision makers at all levels of management. Yet, there is more than sufficient evidence to substantiate the impression that, regardless of their objectives, the use of models by high-level decision or policy makers in the public sector is quite limited. In fact, in our view, the gap between the potential development of strategic models and policy maker needs and uses seems to be, in some areas, widening.

Purpose

In this chapter we describe this phenomenon—largely from the point of view of the policy maker—and attempt to clarify some of the misconceptions and related issues. Insofar as the literature contains previous work and positions on this issue, we acknowledge those who have contributed in the past. However, we feel that by and large the modeling profession, as is the case in varying levels of degree in other professions, has been somewhat reluctant to recognize and criticize itself for these kind of shortcomings. It is our thesis that model developers, model users and decision makers must collectively seek an improved process of development and eventual use of models for policy or strategic purposes. We must learn from the past and overcome the reasons which have led to over-criticism and a reluctance to use models to assist directly in policy making.

Scope

A principle of effective writing is that one should orient his message toward a specific audience. The reason for this rule is obvious, as the background and needs of the intended audience constrain and define what can be read and absorbed in a final product.

There is some question as to whether a similar match is sought with the same rigor in the modeling community. In the early days of model building, the specialists in social science were only accidentally the ones who were interested in empirical model design. Because of this and other reasons more closely related to the existing Gestalt of the mathematical community, modeling grew with an emphasis more on technique than on analysis. To a large extent, this remains as a problem.

In the following pages we begin by defining what we mean by the policy maker and policy making (in modeling terms what we hope are the "users"). We recognize the difficulties in separating policy making from decision making; policy from action; management actions from staff recommendations, etc. However, from the point of view of the policy maker, this difficulty may be academic. In fact, this issue may well be the fundamental basis for the problem we define and describe throughout this chapter.

Next, we discuss policy-oriented (or strategic) models—what they purport to be, what they are not, and what in our view they ought to be. Further, we approach the "needs" of policy makers, attempting to structure a process wherein these problems may be minimized or even eliminated.

Finally, we conclude with a discussion of opportunities for model builders and policy makers because we believe though the challenge is great the payoff is substantial. Policies and policy-level

Preceding page blank

decisions have been made largely on the basis of intuition, staff judgments and default. If the addition of the use of models can really improve the process, then the opportunities should be identified, discussed and acted upon without delay by all who seek to advance the profession and to improve public policy.

Also included as an Appendix to this chapter is a relatively lengthy description of the Strategic Environmental Assessment System currently being developed within the Environmental Protection Agency. This discussion is meant to exemplify some of the many ways strategic models can be designed to be of more direct use of the policy level.

We feel it important to emphasize at the outset that, by the modeling profession, some of the views expressed in this chapter may be interpreted as almost "ascientific" in the sense that we attempt to describe a policy need. As the chapter indicates, such a description of policy making and the opportunities for models to assist in the process does not follow the standard approach of defining the problem, testing alternatives, etc. In fact, the process is probably more subjective and qualitative than scientists would hope. The significant point is that policy-level problems do not always lend themselves to scientific analysis—much less to definition. Thus, the means by which they can be addressed are not always as clear as the kinds of decision problems we find addressed in the literature and "solved" by models. This is not meant to imply that it is therefore impossible to construct models for assisting policy making but, rather, that a new perspective is necessary in order to modify the process by which models are developed, tested, validated and applied in order for them to be helpful to the policy maker.

II. POLICY MAKING DEFINED

Within the scope of this *Guide*, the term "policy making" is restricted to the public sector. At this level of description, the policy maker will be recognized as the person or group of persons (e.g., his immediate staff) who make the principal decisions which guide the public sector. Within the institutional form, these persons are elected (or appointed and confirmed) by a constituency whose desires they presumably (and normally) serve to operationalize and, in fact, they are responsible to the constituency for their actions.

To accomplish this purpose, these policy makers operate with a staff which is normally divided along functional lines and is specialized in carrying out specific portions of the policy program. Figure 1 illustrates in simplified fashion the paradigm of the policy maker. It also points out one of the anomalies of the departmental system which helps explain some of the difficulties involved in modeling for policy uses. Although one description of the policy maker's role in this nation is that of a translator for implementing the public will, the existence and roles of the media and professional societies (as spokesman for the public) define a direct link which often bypasses the policy level. It is the existence of this "shunt route" which lies at the root of one of the basic problems of modeling for policy makers; namely the professional requirement to satisfy not only the needs of the policy maker, but also the demands of the professional community. Because policy makers change as a function of elections and one's professional career is predicated on peer-group status, it is usually the case that the modeling community (and staff for that matter) often tends to be more loyal to the latter than to the former.

As a result of the nature of policy and the peculiar characteristic of departmentalization described above, it is particularly difficult to define "policy making" for the purposes of advocating the use of models for assisting in the process. It is possible, however, to generally bound the process by describing policy making in terms of levels and scope.

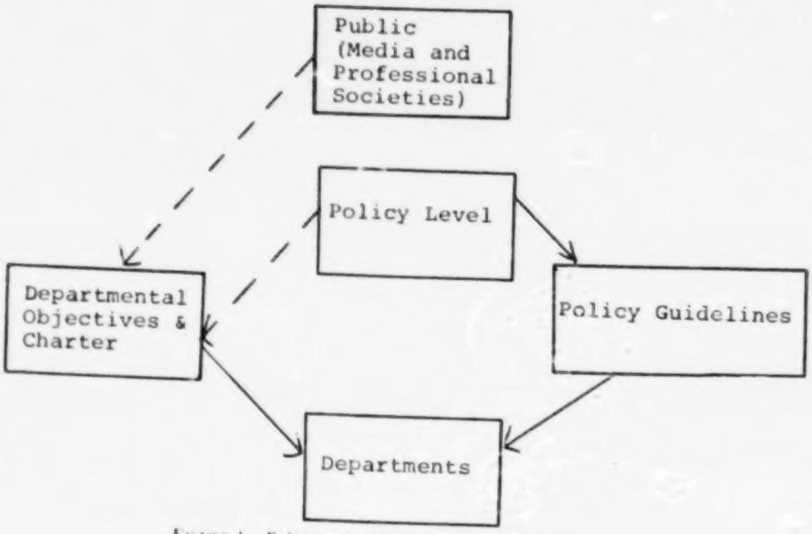

Public (Media and Professional Societies)
Policy Level
Departmental Objectives & Charter
Policy Guidelines
Departments

FIGURE 1—*Policy and the Organizational Structure*

Levels and Scope of Policy Making

It is important to acknowledge the obvious differences in levels wherein public policy is made and to illustrate the differences in scope of problems. In recognizing these differences we mention the levels of policy making and point out the similarities in process such that the principles suggested in this chapter can apply to all policy making that fits our description. It should also be recognized that such a discussion is, of necessity, oversimplified. However, since we are not specifying particular policy problems or case studies in this chapter, generalizations should suffice for illustrations.

Our hypothesis then is that policy making can be defined as those actions (or considerations) taken by elected officials, or their immediate staffs, at all levels of government; and that the process of policy making at this highest level is differentiated from that at the staff, or even program level, due to the scope of problems, issues and necessary decisions that must be made on behalf of society. The following descriptions of levels serve to illustrate this idea.

At the Federal level, the President, his immediate cabinet, the White House staffs, the heads of Executive Departments, the sub-heads, and Office Directors are all policy makers but with very differing problems, needs for information, analytical requirements, and so on. Then too, each Administration differs in its centralization/decentralization of policy making and consequently each "job" varies in its authority and responsibilities.

The making of Federal policy is an interesting process and one that has been well discussed in the literature. For purposes of illustration, we describe below a common problem—urban policy making—and "trace" it through each level of government in order to describe the differences in scope and ramifications. Certain aspects of urban policy have been, as one of the chapters in this book discussed, amenable to the use of models.

The Federal government does, in a sense, have a macro-urban policy, although the quality and comprehensiveness of the policy is, and should be, continually criticized. At the highest level, the President reports to Congress periodically on the status of the national urban policy [1]. Essentially, the policy problem at the Federal level is what programs to pursue and in which form to pursue them [2]. The emphasis on the new Federalism with revenue sharing, Federal reorganization, and decentralization of decision making, represents a type of national urban policy. Yet, the policy decision to pursue such courses of action rests on an evaluation of urban programs and their cost-effectiveness. Urban renewal, transportation, education, housing, antipoverty, and pollution control are examples of such programs. Such a policy is normally affected by incremental, yet critically important, decisions on programs. In addition, national economic and employment policies have ramifications for urban policy. An interesting discussion of these matters can also be found in [3].

The important point is that, at the Federal level, policy makers decide on programs designed to meet certain objectives. In so doing, they select from alternative choices and make tradeoffs among options. The process is not obvious, but the results are observable. For example, one need only analyze the President's budget messages to detect the results of such policy deliberations.

There are, of course, different scopes of national urban policy making among Federal officials. Executive Departments, for example, recommend overall macrourban policy while implementing only that related to their area of responsibility. Nevertheless, policy choices are made in a continuing manner in Washington which determine what programs are to be funded (and which are to be eliminated or reduced) at the Federal level for meeting urban objectives.

The significant aspect of the Federal policy process is that sub-levels below the President exist which have implications for scope as discussed below; however, the essential principle is that policy making implies a choice (or choices) among or between alternative program options or, at a slightly lower level, a choice of tactical options within a program.

At the State level a Governor, surrounded by key department heads and advisors, results in a policy making process not too different from that in Washington except for the obvious fewer number of people and smaller budget involved. As with Congress, States have legislative bodies which also make policy in varying degrees and with varying interfaces with the executive branch.

With respect to urban policy, the States have a role similar to Washington, except that the problem also involves choices which seek to maximize the effect on urban objectives by complementing Federal programs with a particular set of programs at the State level. Thus, one important policy issue is what set of urban programs should the State employ to achieve the maximum effect given the Federal programs and how should these programs be administered.

This is not to imply that the States have only to complement Federal urban programs. On the contrary, urban problems are of key importance to many States and their forward, progressive policies have often led Washington. An interesting urban policy issue at the State level is what programs to pursue, given its own peculiar set of tax bases, urban-rural population splits, growth trends, income variance, environmental goals, institutional arrangements, and the like. Although the nature of urban policy indeed varies from State to State, the process is characterized by many similarities.

At the metropolitan level we generally find Councils of Governments (COG) with elected or appointed policy makers normally attempting to act, in the absence of metropolitan government, in a coordinated fashion to achieve whatever mutual goals are identified. COGs are not, in themselves, elected governments. There is significant variance in COG structures and policy roles throughout the nation, but fundamentally the policy making level represents a coordinative approach to metropolitan-wide problems such as transportation, housing, crime, environment, etc.

COG's are almost universally supported by Federal policy, consequently urban policy issues are viewed in much the same context, albeit at a more micro-level. The fundamental role of COG's is to establish community goals and actions in a coordinative manner, seeking local elected-official support and participation. Thus the metropolitan policy maker responsibility is perhaps characterized more by the need for comprehensive planning than the other levels. It is not surprising, then, that most urban models have been developed for metropolitan problems. Models, however, have not been particularly influential on metropolitan policy, which helps to substantiate the essential thesis of this chapter. A good illustration of this problem at the metropolitan level is contained in [4].

At the local level, whether it be a county with a chief executive or commissioners, or a city with a mayor or council, the level of policy making is fundamentally that of micro-analysis and "firing line" actions. The process, however, is generally the same as that above in that the policy maker relies on a key immediate staff for recommendations and, to the extent he can, makes policy on the basis of intuition, ability to make tradeoffs or by default.

Urban policy-making at the lowest level of government, consistent with the "firing line" analogy above, is characterized by tactical issues, although budgetary problems can be of strategic nature. For example, the choice of budget amounts for programs (e.g., streets, fire protection, education, etc.), represents, whether consciously or not, a strategic set of priority decisions which reflect urban policy. The fact that the budgeting process is accomplished more and more often in the political arena is illustrative of the public realization that financial allocations do help translate policy into action. Yet all too often policy makers (in this case local elected officials) are not fully prepared to respond to the public in this new open fashion.

It is this characteristic that can be found throughout the policy-making process, regardless of the level under study; namely, the consistent need to decide on a course of action (or, significantly, to take no policy action) without sufficient information or time to consider all the ramifications of the possible decisions. It is also this characteristic that makes policy making in the public sector related to, although not equivalent to, management decision-making in industry, and we attempt to describe this analogy below.

Scope of Problems

Concomitant with the highest level of policy making is a set of problems which is often ignored by model builders, or even political scientists and public administrators, except in a theoretical sense. Yet political figures make critical decisions on directions of courses of actions which involve huge investments of funds and time. The breadth of these issues is often overlooked because of the nature of the political process, but positions are taken and policies are made anyway.

For illustrative and comparative purposes, consider briefly one of the uses of models in industry. The overall objective of any industrial concern is to maximize profit. (We ignore for purposes of this discussion more sophisticated measures of commercial effectiveness such as return on investment, market share, etc.) This desire is shared by the president of the company, by the board of directors who elect him, and by the stockholders who elect them. Provided they are receiving what they consider their fair share, it is also shared by the labor force. This means that it is possible to state an overall goal or an "objective function" which is generally consistent for all parts of the concern. Further, this goal is part of the ethos of the country and is legitimized in the law.

Of course, this goal cannot be sought after exclusively, nor can it be sufficient to represent the total business affairs. The side conditions set by other industries, supply and demand conditions, and customer opinion serve to temper it. And attitudes, productivity, and many intangibles affect it. Even so, the overriding goal of industry is relatively explicit and the measures of effectiveness are fairly well understood.

Consider the socio-political area for contrast. Regardless of what may be said of business-economic objectives, there is no question that they are more defined and accepted than the possible goals available to the political decision maker. The personal goal of the president of a corporation, in oversimplified terms, is to remain the president. He generally does this by making a profit and by maximizing the growth of his company, over both the long and short run.

An elected (or appointed) official also has the personal goal of remaining in office, but it is less clear what he must do to achieve this. To remain in office he must balance his goal of staying in office with the values and desires of his constituency. Goals of the constituency are often diverse and at times contradictory; yet they are the group from which all his power legally derives.

Moreover, the prospective elected official, to be elected into office, must play adversary politics and thus continually attempt to shift the objectives of his constituency and convince them that he can maximize the new goals. If the two goals (that of the official in office and of his constituency) are perceived to be close to equal, then he will likely be re-elected to office. If the contender can enlargen the gap, then it may be enough to force the incumbent out of office. It is recognized that this analysis is sensitive to time and is more valid the closer it is to election time. It is also likely to be sensitive to the size of the constituency; the strength of the relationship being inversely related to size.

Certainly it would be a misstatement to assert that all policy models should be aimed at maximizing the vote-getting potential of the politician. However, by conceptualizing the politician's duty as satisfying a constituency, rather than by a more readily definable discipline or functional standard (such as efficiency, or cost-benefit, etc.), it may help modify the model design from rigidity in the area of normative solutions, and may begin to structure designs which are broader in scope though admittedly at present, less quantifiable in nature.

This oversimplified yet illustrative analogy enables one to describe the scope of problems with which the elected policy maker is concerned. The key variable—values and desires of the constituency—must be the continual basis for public policy analysis.

There is another perspective by which the policy problem can be analyzed. Most models have been developed to solve problems of a physical nature. The related necessary attention to detail in solving such problems has been one criterion for the use of computers, to free "man" from mechanical, or routinized detail, so that he may turn his attention to questions of "why". These problem-solving techniques are operational in design-optimizations, least-cost solutions, queing service, etc.; i.e., finding best solutions to "tactical" problems created by the selection of a set of policies by a higher authority. Figure 2 illustrates the problem scope in such a context. Analytical support at the policy-choice level is necessarily macro in its attention to detail, providing strategic-type assistance in "direction" rather than in administration. Support to the programs themselves, even though it may be "policy" in nature, should be in more detail at the intermediate level of Program Policy Decisions, and in as much detail as possible for Program Management Problems. In fact, this analysis could be carried deeper and could show, for example, that models that are large, complex simulations are often so engrossed in micro-level detail they are no longer solving problems, but rather are self-generating work. Many functional models are of this type.

Policy making, then, varies by level and by problem scope. Decisions on

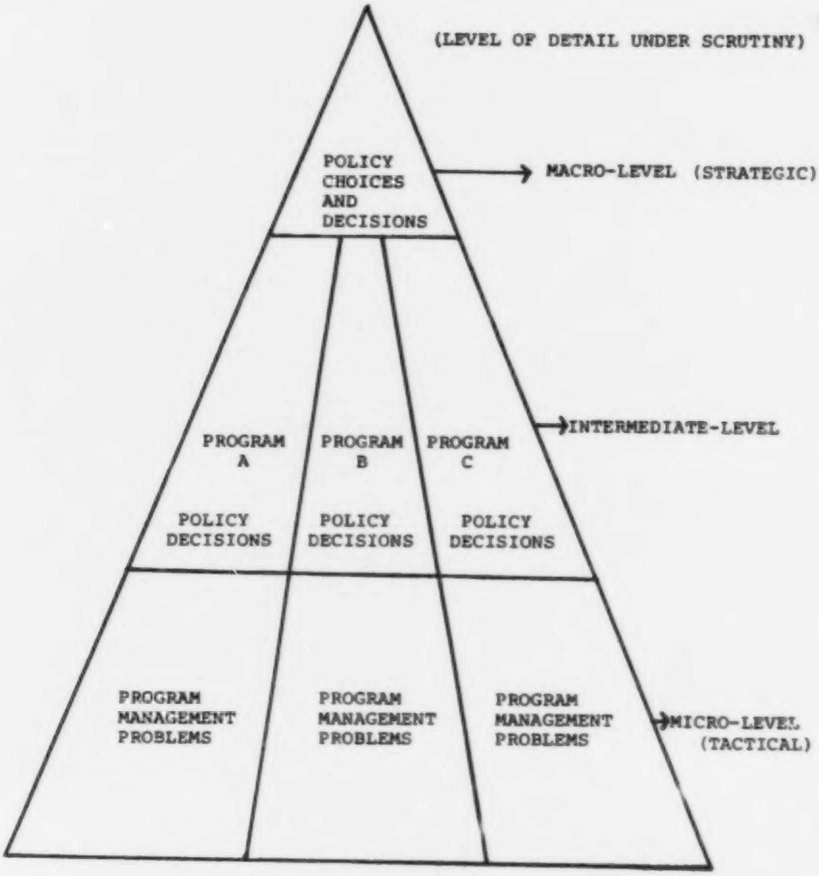

(LEVEL OF DETAIL UNDER SCRUTINY)
POLICY CHOICES AND DECISIONS
MACRO-LEVEL (STRATEGIC)
INTERMEDIATE-LEVEL
PROGRAM A
PROGRAM B
PROGRAM C
POLICY DECISIONS
POLICY DECISIONS
POLICY DECISIONS
PROGRAM MANAGEMENT PROBLEMS
PROGRAM MANAGEMENT PROBLEMS
PROGRAM MANAGEMENT PROBLEMS
MICRO-LEVEL (TACTICAL)

FIGURE 2—*Scope of Policy Problems*

public policy, however, are made continually without regard to such analytical structure. To the extent models can be helpful to the process, and we believe they can given certain conditions, such analytical tools should be developed and applied as an integral part of the process.

III. POLICY-ORIENTED MODELS

Just what is a policy model? If one were to attend a number of the professional meetings held each year or read the articles written by practitioners of the modeling art, it would become apparent that almost all feel that their creations are "policy" models. In fact, the great majority of the models presented in this book were claimed at one time or another to be of this type. From one point of view, many of these models are probably related to policy decisions inasmuch as they do have an input into an overall policy scheme. To leave "policy" at this vague point, however, equates it to the term "decision".

Logically all models can be handled as a subset of decision models, which is the approach described throughout the other chapters. Further, the methodologies used to address policy questions can be viewed as synonomous with those used for any others. It is the contention of this chapter that this logic set is the principal reason for the dearth of models in this area. Although the techniques may indeed be similar or analogous, the problems themselves are so unique that many analysts who propose to build "policy models" appear to fall short of the goal. Therefore, the problem can be said to be one not of technique necessarily, but of content. It is possible that one may find that, having understood the problems, he might wish to develop or expand techniques, but such steps are presently only conjecture. What is worse is that, consistent with our analogy of the businessman above, the models purported to be of policy scope are being used, if at all, to assist in the optimization of the administration of a program or a policy after it has been decided to implement it.

Most decisions to fund model development (or model application) in the governmental sector are made by people who make, or help make, policy. Consequently, it pays for those who wish to build models to profess an ability and willingness for making models useful to such people.

An Assistant Secretary of Transportation once stated this problem in quite a lucid fashion [5]. His essential point was that systems analysis, with its "bag" of models, causes policy makers considerable frustration in that analysts constantly demand time to develop, complete and test a variety of models before policy choices should be made. Yet policy decisions arise and must often be made immediately. Utility of models is rarely the principal goal of the analysts.

In reality, researchers do not interact with the policy maker himself, and only to some limited extent with his staff or with a particular department. Their goals and desires may not agree with those of the policy level, and the resulting lack of specificity relegates models to art forms which merely guess at what is "really needed" by the client. A typical process is that once the contract or work statement is approved, the "client" is no longer the policy maker but one of the staff assigned to monitor the effort. The staff monitor may indeed be quite competent technically, but the policy maker loses interest if his own needs are omitted. This process, having been repeated time and time again in government, has helped cause staffs to become cynical and to seriously question the utility of all models.

For purposes of the policy-model designer, the most important basic goals of the elected official would seem to be the following: first, to satisfy the needs, interests and goals of his constituency; second, to satisfy his own personal needs in terms of pursuing the programs and ideals which motivate his political life, and; third, to be able to select, from a full range of possible policy decisions, those which are most effective in meeting the above two needs.

Within this context, the ultimate policy-making model might, for instance, allow the policy maker to project the effectiveness of various policy alternatives based upon their ability to achieve both his personal goals and their potential effects upon societal values and thus public opinion. The official who understands which of the issues are really the most relevant and which should be given the greatest weight will maximize his vote-getting potential. And the official who can weigh beforehand various policy alternatives in terms of their efficacy toward achieving societal goals will be able to make more effective decisions, and be able to judge their developing impact more accurately in terms of his original goals.

In short, a policy model should assist the policy maker in making policy choices more effectively according to a certain set of clearly identifiable goals. This helps to explain why so many of the planning models available today are unsatisfactory at the policy level. Such models have often taken several years to develop and are likely to be irrelevant to the contemporary goals of the politician and his constituency.

Based on the above considerations, the research direction should be towards discovering the basic aspirations and strategies of the currently elected officials. This suggests, possibly, a greater emphasis on voting behavior, political structure, and personalized goal orientations than are contained in most present models. It may also signify the need for more explicit statements in policy models of the overall societal implications of possible policies. An example of this point is given in [6].

IV. NEEDS OF POLICY MAKERS

One of the basic problems associated with policy modeling is the difficulty in developing a clear definition of user needs, agreed upon by both the model builder and the policy maker. Generally speaking, there appears to be little interaction between the policy maker and the model builder during the initial and design periods of the model. To the contrary, there is often a feeling on the part of the model builder that he understands the problem, and certainly has mastered the analytical technique to be used; consequently, there follows typically much discussion in the way of briefing or educating the policy maker as to what is "good" for him. Seldom, however, are there any meetings in which the focus is inverted and the designer himself is educated. Little wonder, therefore, that the potential users of these models often feel alienated toward them in that the resulting products are not really tailored to their needs.

Interviews and Questionnaires

One technique for establishing such an inverted dialogue, and one with which we are currently pursuing within the Environmental Protection Agency, involves the use of questionnaires in a highly personalized and ongoing interviewing process. The purpose here is to design a strategic policy assessment system (described in the Appendix of this Chapter) which is tailored to the strategic-type needs of the agency administrators. Our hope is that the policy makers will be able to utilize our modeling skills to push the state-of-the-art in the directions of their needs, where such directions might be contrary to our instincts and training.

To begin this process (which we expect to be a lengthy one), the questionnaire discussed below was constructed. In many instances the approach is quite naive, but it was useful in bounding the discussion. We repeat most of the questions here and, rather than report the outcome of our meetings (which are of internal use), we discuss the rationale behind the queries and what we are learning in a general sense from the process.

Questions About the Policy Maker

1. Have you had any training or education in the use of a computer? If yes, what type?
2. What is your personal opinion of the use (or misuse) of computers in policy situations?
3. Have you ever had an experience where you made a decision with the assistance of a model? If so, what do you feel should have been improved in the process?
4. Do you have any analytical training or education? What kind and at what level? From what disciplinary point of view? (i.e., engineering, social science, natural science, etc.). Do you apply this skill in typical policy situations?

The overall purpose of these initial questions is to identify the frame of reference of a policy maker who would be using the model results. Too often, models have little reference to the vocabulary and training of the potential user, particularly if he is to be a policy maker. We recall that several years ago, for example, a study done of the liberal arts, legal, and social science professions noted that usually only economists were really trained to interpret graphics. Consequently, the presentation of information in the form of histograms and graphs added little illumination to the results and, in fact, it was suggested that they added a degree of confusion.

Furthermore, we have experienced an era in which we tried to do something about the real need for large masses of data for decision making purposes. Computer packages were developed which provided manipulative capabilities such as statistical routines and canned analytical algorithms. These packages, however, were generally well beyond the capability of most policy makers. Thus, the policy maker was informed that information was kept in some standard form and that he was able to carry out all sorts of sophisticated statistical operations on it. His training, however, might not actually go beyond high school algebra. To complicate matters, most policy makers have not experienced first-hand an interactive man-machine system. They therefore have no idea what is expected of them or what to expect of the system in such a context.

These points serve to illustrate reasons why the literature and the industry is filled with so-called "management information systems" which most policy makers avoid. To the extent that the policy maker in question has no analytical training, suggesting to him that he personally use a computer or model may be ego-threatening. Secondly, if he has no computer training or understanding, he is apt to expect to be able to do a great deal more or a great deal less with the operation that is reasonably feasible. If, on the other hand, he does possess some degree of training, education, and/or experience is the use of computers in management, an opportunity exists for meaningful dialogue on the capabilities and limitations of models with respect to his real problems.

In short, an understanding of the background and training of the policy maker may tell the model designer what kind of assumptions he can make in terms of input and output to the model system. It will also tell him how much work he will have to do to preprocess data (or outputs) for the policy maker's use. In fact, given the demanding schedule of most policy makers, it is unrealistic to assume that he will have the time to devote to a significant re-education along the line of his analytical or technological capabilities.

Questions on the Policy-Making Process

1. In the context of your role in this organization, how do you best characterize your responsibilities in policy making terms?

2. How do you usually make a policy decision? Do you utilize a staff of advisors, a few principal ones, or decide mostly on your own?

3. How would you like to see policy making done ideally? Why?

4. How and in what form do you prefer to receive policy oriented information—by memo? By report? By problem statements with alternatives? Why?

It appears from reading the preambles to many so-called policy models, or the documentation on same, that few model builders have a clear understanding of how policy is really made. In fact, it may be the case that nobody really understands well this very complex process. On the other hand, it is imperative that the model builder understand not how decision making is done in some abstract fashion, but how the particular policy maker that he is interested in building a model for conceives of and carries out his policy making duties and the specific requirements and resources of the policy maker's role. Likely, this very complex and individualistic process is little understood by the policy maker simply because he carries it out almost intuitively.

The use of computer models, however, requires something more than pure intuition—although just knowing what the policy maker means by "intuition" is in itself useful. Nevertheless, the process must be reduced to explicit form if the model is to have utility. Interestingly enough, discussion of this set of questions will probably provide the greatest education for many model builders who presume to understand the policy level.

Questions on the Use of Models in Policy Making

1. Models (or systems) can be conceived of as answer-generating devices which can be used by policy makers to help solve complex problems. They can also be conceived of as no more than another advisor who is consulted on specific types of questions and whose advice is weighted by the policy maker with that of others. Do you feel comfortable with either of these statements? One more than the other? Why?

2. If a "true" policy model were successfully built, how would you want to use it? For periodic reports of future forecasts? For predicting the impact of past decisions? For frequent use in assisting in day-to-day policy decision making? What kinds of outputs would you prefer—charts? graphs? numbers? memos? others?

3. Would you prefer to interact directly with such a model or delegate its use to your immediate staff for their use in formulating policy advice? Why?

4. It is asserted that modern decision making should take full advantage of sophisticated information systems and hardware. Such a viewpoint suggests either (a) a specially designed "decision room" with full information displays; and/or (b) immediate access. What are your views?

5. Do you face any policy questions on a somewhat continuing basis which a specific policy model might be able to assist in answering?

6. What kinds of questions might other high-level administrators face that are different from those you mentioned? How about the regional administrators?

7. Would you prefer a model which would provide quick summaries of real-world data, or would you prefer a system that enables you to project the probable impact of various policy choices over time?

The above set of questions is aimed at trying to understand, before any real modeling interaction occurs, what the policy maker would anticipate getting out of a model and how he might expect to use it if he would get a model designed to meet his own needs.

Listening to the policy makers who are having models designed and built for them can be a rather disheartening exercise for model developers. The expectations of policy makers usually far outstrip the current technical state of the art. Early specification of these expectations can help the model builder to know whether he is being expected to push the state of the art of modeling significantly to gather new data, or whether the policy problems can be simplified and handled with available tools. In either case, this understanding will help pinpoint his own work in terms of the time, manpower, and resources available.

Further and most importantly, these questions are aimed at exactly how the policy maker perceives using the models when done. There are obvious but important options for use which ought to be discussed, if only in a general sense, at the outset. Some of these are as follows:

a. use in a "situation room", with continual use for policy purposes;

b. use intermittently for "status of the system" reports;

c. use personally, by interacting with a terminal or video display unit;

d. use by policy and support staffs; and

e. combinations of these.

V. OPPORTUNITIES FOR MODELLERS AND POLICY MAKERS

This far we have described the policy-making process, policy-oriented models, and one approach to the determination of the needs of policy makers, in largely a critical context. In this section we attempt to structure in a general sense the kinds of challenges we envision the modeling community and the policy makers must address in order to take advantage of the significant opportunities for improvements in public-policy making that effective computer models can provide. A detailed and quite interesting discussion of "meeting the challenge" is contained in [6].

Challenges to Modelers

There are several schools of thoughts as to what types of models, not only in form but in content, would be of most use to the policy maker. As with all problems, at various levels of abstraction various alternatives are relevant. For instance, if a specific public issue reaches significant proportions it will rise to the policy level and, even though in greater detail than generally handled at this level, may require precise inputs and rigorous handling. Perversely, the detail and data requirements of such models, although potentially of great use to the policy maker, generally preclude their application. Problems of this sort often encompass a time constant which coincides only by accident with the time constant of the scientific community manipulating the models.

More to the point, there seems to be a class of problems which the policy maker deals with all the time, and to which the modeling community has only recently begun to address itself. The policy maker, like the biblical Solomon, continually must adjudicate between departments with competing needs for scarce resources. He will tend to prefer those policies which appear to be most effective in pursuing his goals and in satisfying his constituency and, at the same time, minimize the negative feedback from those parts of the system which he has been forced to de-emphasize. Such models are the type that trade off various policy choices among comprehensive alternatives. In fact, such tradeoffs have to be accomplished within the boundaries not only of relating all the parts of his interest area to each other, but also of relating them through time, so that a short-run decision which maximizes an immediate gain does not turn out to be suboptimal in the long run. Consequently, what appears to be called for is a model which is at a gross or aggregated level of detail, allows manipulation of a large number of alternative choices, is sensitive to the time dimension, and is relatively easy and quick to load, run, and interpret the output in policy terms.

If it is true that the policy maker really could make use of such models, why is it that they have not generally been built or applied? A perusal of the literature provides three general answers to this question: (1) theories available to build such models, particularly those of the social/political institutions, are barely hypotheses; (2) the specific functions constructed which purport to represent the system under consideration are open to question; and (3) the data are either unavailable or in the wrong forms.

One of the largest single issues subsumed in the above is that the proper method of validating policy models is unknown. Since such models may claim completeness, or a comprehensive base, the only test that apparently can be done is to see whether they will replicate one time period. However, if it is truly a comprehensive model, it will only do this on a probabilistic basis. Consequently, the lack of precision generally found in many partial models leaves the analyst with no rigorous method of validation. Since large scale forecasting models can often be used to generate a series of possible alternative futures, which may or may not be likely depending upon the assumptions made as the model cycles through time, the range of outputs would have to evaluated on a "most likely" basis rather than on one which was deterministic.

What this means to modelers may be that since we continue to apply partial models despite the fact that perfect validation of such models is impossible given all the time and detail represented, should we avoid developing policy models for the very same reason? As Forrester discussed in [6], policies made in the framework of complex systems (e.g., urban areas, national governments, the environment) are subject to such interactive behavior in the systems that unexpected and counter-intuitive results may well be the rule rather than the exception. With this phenomenon in mind, how can we rationalize complex partial models in the absence of comprehensive policy-making processes? It is clear that more attempts must be made to model systems and that lack of validation cannot by itself be allowed to obviate policy models. On the other hand, this statement should not be mistaken for license. Some test of reasonableness will have to be made and the standards of scientific integrity kept high. In short, the plea is made for further experimentation and limited use as an alternative to the presently fashionable ostracism. The significant principle should be that specification of goals for the system should be explicit and objective [8].

The question of unavailable data is interesting. Probably little frustrates the policy maker who supports a large organization more than to be told that there are not enough data available to answer his questions. This reply, in fairness to both sides, is probably both right and wrong. From the point of view of the researcher or analyst, the policy maker has a habit of changing his mind as to what information he wants and thus changing his long-range information needs. The time frame is often such that just as the research community begins to gear up and handle questions of one type, the policy maker, having made his mind up in that area, is anxious to move on to other questions. This means that the scientifically-trained analyst is eternally frustrated as he attempts to get the ever-finer detailed information. This is particularly true since, again, his judges are his peers and not necessarily the policy maker.

Nevertheless, it does the policy maker little good to be told that in two years a finished model can give him an answer, when he needs an answer today [5]. From the point of view of the policy maker, his responsibility is to his goals and to those of his constituency. The issues and questions that need answers are not necessarily those that are amenable to the scientific method. Issues and needs change more rapidly than do systems for data collection. Pressures are such that answers must be given, and policy choices made, whether or not enough data are available.

Thus the challenges to modelers are substantial. If we really want policy makers to use our tools for improving public policy, then it appears the standards must be modified, as should the ways of going about developing models and communicating and sharing them with policy makers. And the process should be opened to more involvement on the part of policy makers (and to the public too, for that matter).

But who constitutes the modeling community? In the first place, it is a misnomer to label model builders a "community". If such a community existed, there would be little function to this book. In truth, when modelers from various disciplines assemble, their interaction problems are similar to those of getting any multi-disciplinary team to work together. The journals they read, the training they have, the vocabulary (not only that of the particular discipline, but also that of the mathematics and computer languages) are all different. Most ad hoc groups gathered together to support the policy level spend an inordinate amount of time in just communication problems among themselves, not to mention the substantial difficulties they encounter interfacing with the policy level.

In our view there is a need for a group of policy-oriented scientists to support the policy maker. These people would have the responsibility for making available, to the policy maker, timely information on demand which will help him make his policy decisions. The information would be the best available at a particular point in time and the specific requirements of the policy maker may or may not engender future research projects from the general R&D staff. Models would be a basic foundation of such a support group. Of course, we recognize that since the data and structure of such models are still in their infancy, advice to the policy maker as a result of using general purpose models will have to be of a probabilistic nature, giving him some confidence level as to how good or reliable the estimates are. Nonetheless, such timely advice could be of great value to policy level problems.

For an example of such analytical support capability to the policy level, let us again focus on one area—the environment. Conceptually, a policy maker in any area could be supported in essentially the same fashion.

It is implicitly assumed that there must be comprehensive and long-run planning of the environment if policy is to be promulgated which will actually result in both cleaning up and preserving the environment. In practice, however, the day-to-day implementation of policy often does not accommodate such concerns. This feature is obviously not a function of capriciousness of the advisors, but often of the fact that these people are also constrained by training and by attempting to satisfy their professional community. For instance, perusal of college catalogues shows that, for those who majored in the areas of law, social science, business, public administration and liberal arts more than a decade ago, few had any training in analytical methods. This means that a large portion of our policy leaders and their advisors are conversant with modern decision-making tools only to the extent that they have educated themselves. The subsequent fear of the new and unknown is a serious impediment to change.

There does not yet appear to be any group which would be dedicated to supporting the policy maker in environmental issues, as a full-time vocation. To further illustrate this void, visualize the following typology:

Basic Research		Transfer Agent		Transfer Process		Output
Physical & Natural Sciences	→	Engineer	→	Market Place	→	Product
Social Sciences	→	Planning or Operations Research or Popularity	→	Political Arena	→	Policy

In order to facilitate the transference of a select number of research findings, the portions of society concerned with production support a group called "engineers". The engineers are accepted by not only the producers but by the basic researchers, thus engineers develop their own professional ethic. Because this group of professionals is supported by the producers, the objectives of the engineers reflect the needs of this group. The engineer prides himself on being a problem-solver. He seldom looks at problems from the perspective of how objectives can be accomplished [8].

The policy maker in society really needs this same sort of support. Unfortunately, many who would otherwise fulfill this function find it too risky, in a professional sense, to become tainted with "merely" implementing somebody else's research. Those who made the transfer between basic social research and the political arena have sometimes been labeled "popularizers", especially by the basic scientist who does not see the translation of his jargon to more generally readable form as being useful in the sense of advancing the state-of-the-art.

Another group, who are more readily identifiable, have attempted to fill the gap. These are the planners. Their profession has grown up as complementary in that they have tended to define areas of interest in line with political and administrative jurisdictions. Unfortunately for the policy makers, these people are not usually available for day-to-day support, nor do they usually possess the skills necessary to handle the analytical chores required by comprehensive, long-range policy making. Thus this gap is being rapidly filled by still another contingency—the operations researcher or management scientist.

The purpose of this sorté through the search for a legitimized policy scientist in the environmental field, (or environmental analyst), is to suggest that only when such a profession is implemented will the policy maker have the impetus to take the future into consideration. The environmental analyst will have this long-range viewpoint because of his training, and his logic will be reinforced by his peers. The policy maker will have the confidence (rightly or otherwise) to meld these factors into his policies, since his advisors will do it for him, attest to the soundness of the practice, and communicate it well to the policy maker. It is not, in short, logical to assume that the policy maker will change his habit pattern to a less certain strategy. Therefore, the change will have to come from his advisors, who concomitantly will have to receive support and encouragement, either from other social scientists or from their own peer groups.

Therein lies the challenge to modelers.

Challenges to Policy Makers

This treatise would not be complete if it omitted a discussion of the ways in which policy makers can make better use of analytical support to their policy process. This is not to say that models should be used regardless of whether or not they are of immediate value to the process, but rather that the policy maker has a responsibility to ensure that public policy-making is as rational as possible and takes into account the full ramifications of policy decisions—both in the short and long run.

What this means in terms of the development and use of models is that policy makers should insist on being involved in the process. Most effective model builders would be elated to have policy-level scrutiny and input to their design process because it will greatly enhance the utility of the models.

Furthermore, policy makers should take the time to stay generally abreast of the field. Seminars, literature and staff briefings are important inputs to a continuing analytical capability. An intimate knowledge of the responsibilities of the job coupled with a general understanding of analytical tools can produce a credible and rational framework for not only making policy decisions but also guiding research, studies, and other developmental efforts before projects or issues become crises or failures. It is recognized that time is critical to a policy maker; however, the value of effective and relevant seminars has been proven to be significant.

Finally, policy makers should be innovative and willing to experiment with new tools. Just as modelers have a responsibility to develop tools for policy makers, management should be receptive to new approaches. This is an extension of the interviewing dialogue, directly to the use of new tools, their critique, and feedback toward their improvement. Improved public policy will only evolve from a better understanding of policy needs, impacts and implications.

VI. CONCLUSION

In this chapter we have discussed the general problems associated with the use of models at the policy level of government. In the context of this *Guide*, this chapter hopefully provides some insight into the gap which exists between models and the policy maker, and attempts to clarify the policy-making process as differentiated from program decision-making and lower levels of problem-solving.

Model building for policy use continues to be largely an art, with only slight characteristics of science a part of the process. Policy making, on the other hand, is continually subjected to pressures for systematic approaches and analyses—often by the very people who design models. Unless this gap is closed and a better common understanding and appreciation evolves, the many models described in the chapters of this *Guide*—the majority of which have seldom, if ever, been used by the level of policy maker we are concerned with—may represent the ultimate achievement of models in the policy-making process.

References

[1] *Report on National Growth 1972*, President's Report to Congress, 1972.

[2] Moynihan, Daniel P. (ed.), *Toward a National Urban Policy*, Basic Books, Inc., 1970.

[3] Perloff, A. and H. Wingo, *Issues in Urban Economics*, Resources for the Future, Inc., 1969.

[4] Levin, M. R. and Abend, N. A., *Bureaucrats in Collision: Case Studies in Transportation Planning*, The MIT Press, 1971 (especially p. 241).

[5] Remarks by Paul W. Cherington, Assistant Secretary of Transportation for Policy and International Affairs, before the Transportation Research Forum, October, 1969.

[6] Forrester, Jay, *Urban Dynamics*, The MIT Press, 1969.

[7] Brewer, Garry D., *The Politician, the Bureaucrat and the Consultant: A Critique of Urban Problem Solving*, to be published in Spring 1973 by Basic Books, New York.

[8] Hoos, Ida R., "Systems Techniques for Managing Society: A Critique," *Public Administration Review*, March/April 1973, No. 2.

Appendix

AN EXAMPLE: THE STRATEGIC ENVIRONMENTAL ASSESSMENT SYSTEM (SEAS)

To this point in the chapter there has been little suggestion of potential remedies for our list of inadequacies of policy models. The above questionnaire represents one possible step in the direction of change, an attempt to begin to serve the policy level. SEAS, as a model (or system), has the scope necessary to span the total environment and, ineed, takes into consideration numerous factors whose operation affects the policies and success of EPA but over which the Agency may have little, if any, control. This description of SEAS should be considered an example of a policy model under development. It is not meant to be an implication that there are no other such models, but rather, since the authors are involved, the process is designed to alleviate common and typical problems associated with past efforts of this type.

Scope of SEAS

The SEAS system will utilize preexisting demographic, economic, technological projections to forecast environmental conditions both on a national scale and for the ten standard Federal regions. Inputs will be specified at the national or regional level, and outputs will be expressed in terms of national or regional indicators and assessments.

There are obvious conceptual problems in developing aggregate indicator assessments from information that has its major impact at the local level, such as air and water pollution. There are also policy difficulties in attempting to develop solutions to pollution problems with national (i.e., not spatially specific) policies. In spite of these difficulties, nationally aggregated assessments must be made for the sake of policy makers such as the EPA Administrator, Congress, the Executive and others concerned with environmental progress over time and in relation to other national concerns.

SEAS will evolve to be a comprehensive system capable of forecasting potential future environmental pollution problems and evaluating the response of the environment to such factors as changing goals, policies, standards, and resource constraints.

Figure 3 illustrates a generalized schematic of the overall SEAS system. The system is designed with flexibility to allow evolutionary development with continual improvement over time. The basic core system depicted contains eight components interconnected by a series of functional relationships. The eight components are: Change Agents (Inputs), Processes, Stocks, Residuals, Effects, Reactions, Evaluators, and Output. A general definition of each of these follows:

Change Agents (Inputs)

Population, economic, technologic, or other eventful potentials whose change in magnitude, direction, appearance or disappearance may pose significant consequences for environmental management. Population inputs may include changes in total growth, patterns of distribution, social and economic compositions, and value orientations. Economic inputs may include changes in total growth, patterns of distribution, activity mix, social responsibility, work ethic, and multinational orientation. Technologic inputs may include pollution abatement technology changes, improved recycling technology, etc.

Processes

Major activities of an ecological, socio-economic system which respond to the Change Agents. Processes are divided into five sub-modules: extraction, production, distribution, consumption, and disposal. Each of these has a human and non-human component. The human components are those processes which are initiated, sustained, or purposefully affected by man while the non-human components are those processes which will "naturally" occur without human intervention.

Extraction—involves obtaining raw materials for utilization in the production process. Raw materials may result from mining natural resources, restoration, reclamation, recycling, resource recovery, decomposition, etc.;

Production—is defined as a synthesis of raw materials, intermediate products, energy (labor) and capital to yield a potentially consumable product;

Distribution—involves those activities associated with overcoming spatial separation;

Consumption—defines activities which result in the utilization of the output of the other processes; and

Disposal—defines the collection, storage, transportation, absorption and dissipation of waste materials that are not completely utilized by the other processes.

Stocks

The supply of natural or man-produced resources which are effectively available for use by the processes. Tangible quantities can be sorted into three arbitrary categories—depleted, fixed, and variable.

Depleted Stocks—those stocks which

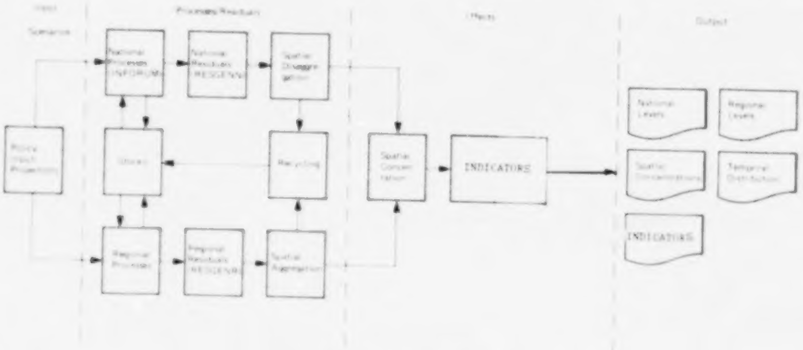

Input
Scenarios
Processes/Residuals
Effects
Output
National Processes (INFORUM)
National Residuals (RESGEN)
Spatial Disaggregation
Policy Input Projections
Stocks
Recycling
Spatial Concentration
INDICATORS
Regional Processes
Regional Residuals (RESGEN)
Spatial Aggregation
National Levels
Regional Levels
Spatial Concentrations
Temporal Distribution
INDICATORS

FIGURE 3—SEAS Schematic

are renewed at a rate so slow as to be considered negligible, i.e., coal, oil.

Fixed Stocks—stocks which have a continuous supply, the process of renewal is not affected by man; and the quantity or quality of the resource may be diminished in local areas, i.e., water, air, etc.

Variable Stocks—the rate of renewal of these stocks depends on the physical environment and magnitude of the propagating or producing stock or process, i.e., agriculture, forestry, etc.

Residuals

Emissions (air), effluents (water) wastes (solid), laydown (pesticides) and releases (radionuclides) that result before controls (gross residuals) or enter the environment (net residuals) after abatement or recycling.

Effects

Residual levels in the different media impact upon the socio-political, economic, and environmental systems. The magnitude and characteristics of this impact will depend upon the deleterious or beneficial qualities of the residuals, the distribution or concentration of the population at risk and the type and sensitivity of the physical environment. These impacts will be considered in the effects modules of the SEAS system.

Reactions

Certain changes in socio-political, economic, and environmental systems occur in response to exposure to residuals as well as actions taken to counteract their impacts. For example, provision of health care facilities may increase, or populations and industries induced to migrate.

Evaluators

Techniques which apply to the analysis of the changes in the Stocks, Residuals, Effects and Reaction Sectors.

Stocks—measurement of the effective growth and decay rates, levels and usage rates of the stocks by the different processes and residuals.

Residuals—measurements of levels and/or rates of increase (decrease) of residuals on a regional and national level as well as a comparison of previous years' levels and rates of change.

Effects-Reactions—analysis of the severity, duration, and dispersion of the effects on the socio-political, economic, and environmental systems and their subsequent reactions.

Output—The output of SEAS will be designed and formatted to provide quick summary options for policy assistance and more detailed outputs for investigative purposes. The output will include graphical displays and formatted statistics by region and a national summary.

This brief overview of the model should serve as an introduction to the SEAS concept. It is certainly not sufficiently detailed to satisfy the requirements of a systems analyst, but such documentation is available. For illustrative purposes we use the design to discuss its potential use within the Agency. Because the System is only recently operative, it will be best to discuss it in generalities.

The first important aspect is the scope of the model. The system will attempt to look at all pollution media at the same level of aggregation or analysis. This is a design decision meant to enable the policy levels of the Agency to consider comprehensive decisions, across the Assistant Administrators' responsibilities and lines of authority. SEAS is also to be used, hopefully, to assist the Assistant Administrator for Research and Development to focus on areas of research where information is sparse and where his experience tells him that such information has priority. This means that the level of detail will likely not be satisfactory to some individual users in given functional areas and almost certainly not adequate for the purposes of the scientist-analyst. On the other hand, because the policy maker does not normally get involved in programmatic task level details, the tools used to support him probably should not be at this specificity either.

In addition to the scope of the model in terms of the responsibilities of EPA, the system is sufficiently broad in scope to generate possible alternative futures that might occur without taking any special actions. In the next section we describe the context in which SEAS is planned to be used in accordance with the scope discussed above.

Expected Uses of SEAS

We realize that this description of SEAS is limited; however, it is adequate for the purposes of this paper. There are two specific types of efforts to make the SEAS system most useful to the Agency. In the first place the actual output itself is being designed in close conjunction with the policy level. It will not consist merely of a series of tabular printouts, but will be formatted so that the analysis is consistent, the graphics and tabular information of the type preferred by the users, and in a form with which they are comfortable.

The second output design stage involves emphasis on the type of information that will be generated. SEAS, as a forecasting system, is highly complex and will be able to generate a great variety of information. It is certain that the potential amount of data which could be reported will be huge and, if one were to attempt to report it all after each iteration of analysis, or even a large part of it, the volume may well preclude its use. Therefore, it has been decided that the output will likely be in summary form, in relation to selected scenarios. The scenarios will be carefully defined, again with the help of the policy level, to present possible alternative futures in a form and in the kind of detail that will assist in policy making.

After developing projections based on selected scenarios, SEAS will report on the future state of the environment. The technical staff will, with the help of other experts, make judgments as to which of the alternatives are most likely and then use the other estimates to statistically calculate the probability of the result. Finally, the range of out-

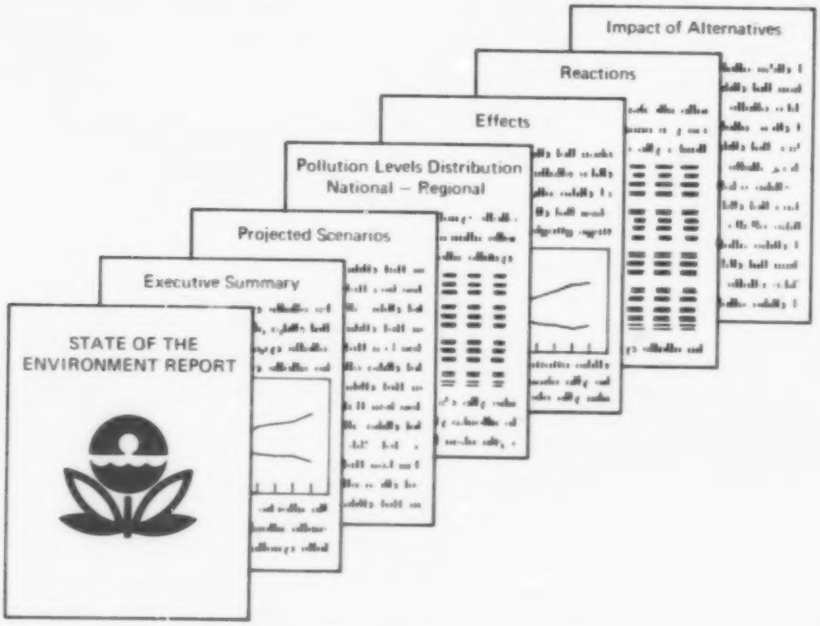


FIGURE 4—State of Environment Report

puts will be presented in terms of "best" and "worst" cases according to the particular interests of the policy makers. Figure 4 illustrates schematically the contents of the "State of the Environment", an annual (or periodic) product of forecasting and analysis of the type described above.

In summary, we have presented SEAS as a system which is to be a tool for use at the policy level of EPA. Its scope accommodates the interests at this level and as such is being designed so as to provide maximum flexibility. The care with which the results of its use can be presented and the attention being paid to making sure that the potential users understood the assumptions in the system, should help mitigate against, on the one hand, unwarranted expectations of the potential of the system, and fear or derision of its real worth on the other. Finally, its policy uses are expected to be broad in scope and, by designing for maximum flexibility, hopefully diverse in nature. The ultimate tests will come when it is fully developed and used.

Chapter 3

Models in Air Pollution

By

Nozer D. Singpurwalla

SUMMARY 63

I. AIR POLLUTION: GENERAL DESCRIPTION OF THE PROBLEM 64

- A. General Description of the Air Pollution Problem 64
- B. The Notation of Averaging Time and Maximum Concentration 66
- C. Decision Processes in Air Pollution 68
- D. Mathematical Models in Air Pollution 69
- E. Relationships Between Models and Decisions 73

II. DISCUSSION OF SPECIFIC MODELS USED IN AIR POLLUTION 76

- A. Stochastic Models 77
  - 1. Models for the Distribution of Pollutant Concentrations 79
  - 2. Regression Models in Air Pollution Analysis 82
  - 3. Models for Predicting Maximum Concentrations 86
  - 4. Time Series Analysis Models 88
- B. Deterministic Models 94
  - 1. Diffusion Models 95
  - 2. Optimization Models 99

REFERENCES 101

Models in Air Pollution

SUMMARY

In this chapter we will discuss some of the mathematical models that can be used to obtain answers to the quantitative questions which arise in air pollution problems. The mathematical models discussed here have proven to be of some value in analyzing specific air pollution decision problems. More important, however, by using these models we are now able to examine these problems more systematically and evaluate proposed policies in more detail.

Since the models that are discussed in this chapter may be too technical for some readers, we have presented our material in two sections. In section I we have given a non-technical discussion of the nature and the use of mathematical models useful in air pollution problems. In section II we have given a more detailed and technical outline of these models.

In section I.A we present a general description of the air pollution problem. We state with a minimum amount of detail, the causes of air pollution in both urban and rural areas, and the ill effects of air pollution to man, animals, plants, and property. We also discuss here the setting of air quality standards and their interpretation.

In section I.B we introduce and discuss the important notion of "averaging times" and "maximum concentration" and point out the relevance of these concepts to the air quality standards. In the interest of brevity and clarity of presentation, this section is written more informally, using mathematical notation.

In sections I.C and I.D, we discuss the various decision processes in air pollution and outline the various mathematical models that can be used in air pollution, respectively. We characterize the models by the functions which they can perform and by the methodology employed to develop them.

In section I.E we discuss the relationships between the models presented in section I.D and the decision processes presented in section I.C. We discuss these relationships by using hypothetical problems faced by various levels of decision makers and how they would use the models to solve their problems. We also illustrate the interrelationships between the various models and decision processes by general diagrams. Readers who are mainly interested in the use of mathematical models and the decision processes for which they can be used will find section I.E a convenient place to stop. Clearly, those readers who wish to gain more insight into these models, their basis and their underlying assumptions are advised to proceed to section II.

Section II is divided into two subsections. In II.A we discuss stochastic models such as regression, time series analysis and the extreme value theory models. In II.B we discuss deterministic models such as the diffusion models and optimization models. Wherever feasible, we illustrate the use of these models by alluding to actual applications for which models have been used. We also discuss the basic assumptions underlying these models, and point out the limitations and the scope of these models.

In summary, this chapter on air pollution models is not intended to be all inclusive and exhaustive. Rather, it is written with the objective of being as complete as necessary in order to convey a basic understanding of those air pollution decision problems for which models can be of assistance.

Preceding page blank

AIR POLLUTION

Models Discussed in Chapter

Summary Table

Model Decision Area	General Type	Important Characteristics
Air Pollution Standard Evaluation, Monitoring Control	Probability Distribution of Pollutant Concentration	Uses data to obtain a probability distribution of pollutant concentrations. This distribution can be used to find out how often a proposed standard will be violated
Air Pollution Standard Evaluation, Monitoring Control	Probability Distribution of Maximum Concentrations	This distribution can be used to find out how often a proposed standard for maximum concentrations will be violated
Prediction of Pollutant Concentrations, Identification of Significant Factors Causing Pollution, and Economic Analyses of Air Pollution	Regression Model	Based on data we can evaluate what factors cause pollution. We can also predict concentration levels based on specified values of the significant factors
Prediction and Forecasting of Pollutant Concentrations, Identification of Trends	Time Series Analysis Models	Using historical data we can forecast individual values of future pollutant concentrations. We can detect trends in pollution levels
Prediction of Pollutant Concentrations at Different Points in an Area	Diffusion Models	Using some data on wind velocity, wind direction, and other meteorological parameters, concentrations at various points in space can be predicted
Optimization of Resources, Planning Strategies, Making Decisions	Linear Programming Nonlinear Programming	Gives us a strategy for optimum allocation of limited resources such as fuels, monitoring equipment, etc. to minimize the control effectiveness of an air pollution program

I. AIR POLLUTION: GENERAL DESCRIPTION OF THE PROBLEM

A. General Description of the Air Pollution Problem

To a decision maker concerned with environmental problems, the subject of air pollution is of interest for compelling reasons. While many aspects of this subject are political and emotional, some of the major questions are quantitative and indeed mathematical. Air pollution comes from many sources, varies widely from place to place, and time to time and affects many individuals differently.

The automobile problem involves problems of traffic flow, and the varied performance of vehicles of similar or disparate models. Atmospheric conditions add major variability to a problem already characterized by variation. Clearly, the building and testing of mathematical models has to be a feature of any attempt to become cognizant of the problem as a whole.

Whenever combustion occurs, air is polluted. Man pollutes his atmosphere by burning fuels to produce energy, by waste incineration, by discharges of gas and aerosol by-products from chemical processing, manufacture, construction and demolition. As the rate of fuel combustion has increased over the years, the emission of pollutants into the atmosphere has also increased. Though the problem of atmospheric pollution is world-wide, it is primarily a problem of municipalities where pollutants are concentrated in relatively small geographical areas that are so densely populated. The geography of a certain area plays an important role in the generation of an air pollution episode for that area. The temperate zone (latitudes 30°–60°) experiences more variability in weather than either the polar regions or the equator. Cold air masses are formed at the poles and hot air masses in the tropics. Variations in these air masses combined with the rotation and the roughness of the earth's surface cause large pieces of these air masses to break off and move through the temperate zone. The transition regions between masses of cold and warm air are called "fronts." An "inversion" is said to be formed when there is a layer of warm air above one of colder air. In such a situation, in the absence of wind, pollutants entered into the atmosphere will not be dispersed, creating a situation of concern. Understandably, air quality is strongly dependent upon weather. Superimposed on the influences of large weather systems are the complications introduced by cities themselves. Any urban complex exerts a notable effect on its own climatology. The many structures rising to different heights and arranged in different patterns, the properties of the materials used in these structures and the pavements that surround them all play a role in creating a localized atmospheric system.

Air pollutants offend and harm man, animals, plants and property in various ways. One offense is odor, an effect that can be sensed if pollutant concentration exceeds the odor threshold for a few seconds. Another effect is surface damage. For instance, sulfur dioxide, an acid pollutant, damages the surfaces of man's upper respiratory tract, plant leaves, buildings and metals. A third effect is the toxicity created by pollutants that enter the lungs and dissolve in the blood and body tissues; carbon monoxide and lead are such pollutants. Usually surfaces and internal tissues require a pollutant exposure of an hour or more before much damage is caused. Carcinogenic pollutants may cause cancer from a lifetime of accumulated exposure.

The setting of air quality standards is a major step toward controlling widespread pollutants. The major purpose of air quality standards is to prevent the occurrence of adverse effects such as those described above.

The concentration of an air pollutant is a measure for expressing the degree of pollution. The concentration is expressed in either parts per million (ppm) or micrograms per cubic meter ($\mu g/m^3$). In the United States, the Federal government sets national minimum air quality standards for pollutants; however, the states may set more stringent standards if they wish. In these standards, allowable levels of air pollution are expressed in terms of the concentration of a specific pollutant and the duration of exposure to that pollutant.

Table 1, abstracted from the "Office of Air Programs Publication No. AP–89" shows the effect of various pollutants as a function of the duration of the exposure. For example, an exposure to sulfur dioxide for a period of 4 days impairs one's general health, whereas an exposure of sulfur dioxide for a period of one year or more would damage vegetation, corrode materials and of course impair one's health. At the other extreme, an exposure to oxidant of one second or more may cause eye irritation. It is apparent from this table that in order to relate pollutant effects to pollutant concentrations, the concentrations should be analyzed as a function of exposure duration.

The mathematical models useful in air pollution analysis refer to pollutant concentrations and the duration of ex-

Table 1. Relationship Between Air Pollutant Exposure and Effects

Approximate exposure duration	Effect	Pollutant			
		Carbon	Oxidant	Particulate matter	Sulfur oxides
1 sec	Sensation Odor Taste Eye irritation				
1 hr	Student athlete performance impaired Visibility reduced				
8 hrs	Judgment impaired Heart patients stressed Vegetation damaged				
1 day	Health impaired Soiling				
4 days	Health impaired				
1 yr	Health impaired Vegetation damaged Corrosion Soiling				

posure, depending on what the model is attempting to do. It is this aspect of the problem that makes mathematical models for air pollution quite involved.

Before we proceed any further, we would like to point out to the reader that the mathematical models used by decision makers or by other analysts dealing with air pollution problems are quite involved and complicated. For example, it is well recognized that diffusion models useful to both decision makers and other analysts are derived from the complex phenomenon of heat transfer and turbulent flow, and are represented by lengthy and complicated equations.

To be able to discuss and motivate the use of these models in any satisfactory and convincing manner, it is imperative that we introduce some mathematical formalization into our discussion. What follows is with this spirit in mind, though we have attempted to minimize the mathematical details to a point of absolute necessity. We urge the interested reader to plow through some of these fine details which are presented in a style and manner palatable to a non-technical reader.

To be able to discuss the mathematical models presented here it is necessary that we acquaint the reader with some notions fundamental to air pollution. This is the notion of an averaging time and maximum concentration. A discussion of these notions also familiarizes the reader with the manner in which air pollution data is collected.

B. The Notation of Averaging Time and Maximum Concentration

In an effort to relate air pollutant concentrations to the duration of exposure, the notion of an averaging time is introduced. Under the Continuous Air Monitoring Program (CAMP) of the Environmental Protection Agency, pollutant concentrations are punched into a tape every five minutes. Let t_1, t_2, ... t_k, ... t_n denote the instants of time, spaced five minutes apart, at which concentrations of a certain pollutant, say $x_{t_1}, x_{t_2}, \ldots, x_{t_n}$ are recorded on a tape. Thus x_{t_i} denotes the concentration (in parts per million or micrograms per cubic meter) of a pollutant observed at time t_i.

Larsen [1] refers to the x_{t_i} as values of the concentration for a five minute averaging time. Values of the concentration α_i for a ten minute averaging time are obtained as

$$\alpha_1 = \frac{x_{t_1} + x_{t_2}}{2}, \alpha_2 = \frac{x_{t_3} + x_{t_4}}{2}, \dots .$$

Similarly, values of the concentration for averaging times of fifteen minutes, say β_i are obtained as

$$\beta_1 = \frac{x_{t_1} + x_{t_2} + x_{t_3}}{3}, \beta_2 = \frac{x_{t_4} + x_{t_5} + x_{t_6}}{3}, \dots .$$

Similarly, one can obtain values of the concentration for averaging times of twenty minutes, one hour, one day, one month, etc.

To be more general, we consider values of the concentration for averaging times of 5k minutes where k is any number, an integer. Thus, averages of length k are

$$k^{-1}(x_{t_1} + x_{t_2} + \dots + x_{t_k}), k^{-1}(x_{t_{k+1}} + x_{t_{k+2}} + \dots + x_{t_{2k}}), \dots ,$$
$$k^{-1}(x_{t_{n-k+1}} + \dots + x_{t_n}). \left(\text{Note: } k^{-1} = \frac{1}{k}\right)$$

For purposes of evaluating air quality, it is important to know the behavior of the maximum average pollutant concentration since it is the maximum concentration lasting for a certain duration of time that causes concern. Let $\eta_{k,n}$ denote

$$\eta_{k,n} = \text{Max}\{k^{-1}(x_{t_1} + \dots + x_{t_k}), \dots , k^{-1}(x_{t_{n-k+1}} + \dots + x_{t_n})\}.$$

The above expression states than $\eta_{k,n}$ is the maximum of the averages of length k, from a total of n observations.

The following table (Table 2), abstracted from Larsen [2] shows the maximum average concentration for certain pollutants adopted as an air quality standard by several states in the U.S.A. and West Germany and the U.S.S.R. It is clear from this table that air pollutant concentration standards are stated in terms of the various averaging times, and it is of interest to know the chances that the maximum concentration exceeds the specified standards.

For example, the maximum eight-hour average concentration of carbon monoxide recommended by New York State's community air quality objectives is 30 ppm, whereas in the U.S.S.R., it is 1 ppm for a 24 hour averaging time. According to CAMP an 8 hour averaging time would correspond to k = 8 × 12 = 96, since one hour would correspond to k = 12. Thus if the maximum of averages of length 96 for carbon monoxide exceeds 30 ppm, there would be a violation of the New York State standard.

As mentioned before, the maximum pollutant concentration is of main concern to those involved with the control of air quality. As a matter of fact, most of the analysis of the CAMP data deals with the maximum concentrations. Table 3 abstracted from Larsen [2] shows the observed maximum concentration of pollutants in some U. S. cities for various averaging times.

For example, the observed maximum concentration of carbon monoxide for Chicago for an averaging time of 8 hours is 35 ppm, a clear violation of the New York standard given in Table 2, whereas for Los Angeles it is 28 ppm, below the New York standard.

The models for predicting maximum concentrations as a function of averaging times using extreme value theory and time series analysis which will be discussed later rely heavily on the concepts laid out here.

The main objective of this section was to introduce the decision maker to the most important variable for his decision making; namely, the standard for pollutants. This standard being expressed in terms of the maximum concentration and an averaging time.

This section, which is a bit technical in nature, has been presented prior to the following section which deals with decision making in air pollution, be-

Table 2. Air Quality Standards

Pollutant	Concentration and Averaging Time for Selected Standards					
	California	Colorado	New York	West Germany	USSR	Other
Gases						
carbon monoxide	30 ppm 8 hr		30 ppm 8 hr		1 ppm 24 hr	5 ppm 8 hr
nitrogen oxides	0.25 ppm 1 hr	0.1 ppm 1 hr		0.5 ppm 1 hr	0.06 ppm 24 hr	
oxidant	0.15 ppm 1 hr	0.1 ppm 1 hr	0.15 ppm 1 hr			0.02 ppm 8 hr
sulfur dioxide	0.3 ppm 8 hr	0.1 ppm 24 hr	0.4 ppm 1 hr	0.2 ppm 1 hr	0.06 ppm 24 hr	
Suspended Particulates						
lead					0.7 $\mu g/m^3$ 24 hr	
total			100 $\mu g/m^3$ 1 yr		150 $\mu g/m^3$ 24 hr	

cause in this section we establish a vehicle for discussing the decision making processes. Had we reversed our order of presentation we would not have been able to pinpoint the problems effectively.

C. *Decision Processes in Air Pollution*

This section and the one which follows essentially capture the essence of decision making problems in air pollution and the role of mathematical models which assist a decision maker. Since the models that are used here are of a highly technical nature, what we have done in this section of Part I and the others which follow is to give a nontechnical discussion of the nature and the use of these models. Part II of this report outlines these models in a greater technical detail and could be pursued by anyone interested in investigating these models in greater depth.

Decision-makers involved with an analysis of air pollution are faced with making decisions which can be broadly classified under the following categories:

- Those involving the setting of air pollution (quality) standards.
- Those involving the monitoring of air pollution concentrations in order to adhere to the specified standards.
- Those involving the control of air pollutant emissions from different point and areas sources in a manner which enables the air pollution standards to be met and at the same time does not hinder the general economic welfare and the day to day needs of the inhabitants of a particular area.

The above decision processes are quantitative in nature, since they involve air pollution standards which are maximum concentrations for specified durations. Mathematical models to help one execute these decisions effectively have been developed. Though some of the models will be specifically applicable to one of the above decision processes, there are some models which can be used for more than one of the above decision processes.

Before presenting a discussion on the role of mathematical models in air pollution analysis, some more details on the decision processes outlined above are next discussed.

The setting of air quality standards is a problem well debated by those involved in air pollution studies. The major purpose of air quality standards is to prevent the occurrence of adverse effects such as those described in Section A.

Once an air quality standard has been set, the next task is to monitor the standard. The question of how often will the standard be violated, is a violation a genuine one or is it due to statistical fluctuations, etc. will be of interest to the decision-maker. The decision-maker will also need techniques which can predict the violation of a standard, and based on this prediction take preventive measures. The decision-maker will also need to know the reliability of his predictions so that he may not raise false alarms too often.

Since pollutants are emitted into the atmosphere through various sources, the decision-maker will often be faced with the problem of deciding which sources of emissions should be controlled. For example, in any urban area, emissions from automobiles, industrial chimneys, individual heating units, power plants, etc. are common. In order to adhere to a specified standard the decision-maker may want to reduce the emission from automobiles by requiring them not to pass through certain areas, or by requiring that power plants shift to low sulfur fuels, or by requiring that individual heating units be closed down. Each of these actions affects the economic well-being and the personal comforts of the citizens in this area. Once again there are conflicting objectives under which a decision-maker is required to act and mathematical models could be useful tools for such decision problems.

These decision problems presented in the above paragraph can be better conveyed by the following examples:

The mayor of a large city must decide whether to approve a proposed major addition to an electric power generating station. If this addition is approved, the residents of this area would be assured their growing demand for electric power at a fair economic cost. However, approval of the addition would lead to a further worsening of the city's air quality. Should this addition be approved?

The governor of a State must decide whether to pass legislation that would place stringent limits on the sulfur content of fuels burned in the State. If passed, this will lead to a definite improvement in the State's air quality but the consumers would have to pay more due to the use of the more expensive low sulfur. Should this legislation be passed?

Each of the above decision problem (or problems) is faced presently or has been faced by public officials. Moreover, they are representative of a host of similar problems that public officials are increasingly being asked to confront, namely, "Should government (State or Federal) adopt a specific, proposed program to improve air quality?" With each, there is the additional question, "What should the air quality be?"

Due to the complex nature of these problems, an individual finds it difficult to decide which, if any of a series of proposed air pollution control programs to support. One cannot claim that mathematical models can provide a solution to any of the current air pollution problems, but using these models one may be able to examine proposed policies in more detail than is currently done.

D. Mathematical Models in Air Pollution

As discussed before, the decision processes in air pollution problems being quantitative in nature, mathematical models play an important role. The variables to be controlled or predicted are the concentrations of the various pollutants for specified averaging times. It is to be emphasized that mathematical models are abstractions of reality and have to be used with caution. Before using these models one has to ascertain that the basic assumptions underlying these models are adhered to, and realize that the conclusions obtained from these models have to be judiciously exercised. In summary, mathematical models are merely an aid, albeit a strong one, in the general decision-making process.

The models that are discussed here can be characterized by the functions which they perform. As far as air pollution problems are concerned, this

Table 3. Maximum Concentration of Various Pollutants in Various Cities for Various Averaging Times

City and Averaging Time	MAXIMUM CONCENTRATION										
	Gases, ppm*							Suspended particulates, μg/m^3			
	Carbon Monoxide	Hydrocarbons	Nitric Oxide	Nitrogen Dioxide	Nitrogen Oxides (NO + NO$_2$)	Oxidant	Sulfur Dioxide	Lead	Organic (Benzene soluble)	Sulfate	Total
Chicago											
5 minutes	64	20	0.97	0.79	1.12	0.17	1.62				
1 hour	46	8	0.91	0.47	1.06	0.13	1.12				
8 hours	35	7	0.49	0.22	0.62	0.09	1.02				
1 day	27	6	0.35	0.15	0.50	0.07	0.79	1.9	39.0	32.8	714
1 week	21	4	0.24	0.10	0.35	0.05	0.45				
1 year	12	3	0.10	0.05	0.15	0.02	0.18	0.6	15.1	16.6	200
Los Angeles											
5 minutes	81		0.97	0.59	1.49	0.53	0.51				
1 hour	43		0.79	0.54	1.39	0.49	0.19				
8 hours	28		0.49	0.37	0.78	0.31	0.09				
1 day	22		0.38	0.20	0.55	0.16	0.07	6.3	69.4	35.9	594
1 week	18		0.23	0.13	0.32	0.07	0.04				
1 year	11		0.08	0.06	0.12	0.04	0.02	1.7	21.3	14.4	185
New Orleans											
5 minutes			0.42	0.14	0.49	0.18	0.31				
1 hour			0.36	0.13	0.43	0.17	0.10				
8 hours			0.15	0.07	0.20	0.11	0.04				
1 day			0.08	0.05	0.12	0.08	0.03	1.0	41.6	18.7	239
1 week			0.04	0.04	0.07	0.06	0.02				
1 year			0.02	0.02	0.04	0.02	0.01	0.3	10.7	9.6	93
New York‡											
5 minutes											
1 hour	27	13	0.42	0.56	0.65§	0.12	2.20				
8 hours	22§					0.08§	1.30§				
1 day	15§				0.22§		1.00§	15.0§	56.0	53.1	532
1 week								9.8			
1 year								4.0§	14.4	22.8	182

Philadelphia											
5 minutes	43	17	1.35	0.37	1.50	0.25	1.00				
1 hour	37	12	1.15	0.26	1.27	0.20	0.84				
8 hours	27	8	0.52	0.14	0.45	0.13	0.54				
1 day	21	6	0.25	0.10	0.28	0.09	0.43				
1 week	17	4	0.12	0.07	0.13	0.06	0.19	4.3	33.6	62.0	411
1 year	7	2	0.04	0.04	0.08	0.02	0.09	0.8	12.2	19.8	177
San Francisco											
5 minutes	47	16	0.68	0.38	0.91	0.29	0.26				
1 hour	43	12	0.66	0.33	0.76	0.17	0.16				
8 hours	42	9	0.52	0.23	0.52	0.10	0.10				
1 day	41	7	0.32	0.13	0.32	0.08	0.08				
1 week	40	4	0.17	0.08	0.23	0.05	0.05		57.2	20.2	253
1 year	7	3	0.09	0.05	0.14	0.02	0.02		9.2	6.2	76
Washington											
5 minutes	44	17	1.65	0.24	1.78	0.19	0.87				
1 hour	41	15	1.28	0.23	1.41	0.18	0.62				
8 hours	34	10	0.54	0.14	0.69	0.11	0.32				
1 day	23	7	0.29	0.10	0.38	0.07	0.22				
1 week	14	5	0.10	0.06	0.15	0.05	0.13		48.6	21.0	268
1 year	7	3	0.03	0.04	0.07	0.02	0.04		13.9	12.4	119

[*] Gaseous pollutant concentrations are for 1 year (1964 for most of the cities) of continuous sampling.

[†] Particulate samples are for 24-hour-duration random samples taken from 1957 to 1961. About 100 organic particulate and total particulate samples were available for each city. About 25 lead and sulfate samples were available.

[‡] Gaseous pollutant and particulate lead concentrations were measured by the New York City Department of Air Pollution Control at the air pollution laboratory, 170 East 121 Street.

[§] Concentration estimated from approximate interrelations among data.

functional characterization leads to two broad categories:

1) Those models that are used for the purpose of predicting or forecasting air pollutant concentrations. The models (to be discussed later) which fall under this category are:
   a) Models for the distribution of pollutant concentrations
   b) Extreme value theory models
   c) Diffusion models
   d) Regression models, and
   e) Time series analysis models
2) Those models that are used for controlling or monitoring air pollutant concentrations, and those models that are useful in understanding the physical phenomena which influence pollutant concentrations. The models which fall in this category are
   a) Diffusion models
   b) Regression models
   c) Optimization models, such as linear and nonlinear programming models

Clearly, there are some models which can serve either one or both of the above two functions.

The mathematical models useful for air pollution can also be categorized by the methodology or the scientific thought employed to develop them. Once again this type of characterization leads to two classes:

1) Stochastic Models—These models have a probabilistic structure imposed upon them. The models which fall under this category are
   a) Models for the distribution of pollutant concentrations
   b) Extreme value theory models
   c) Regression models
   d) Time series analysis models

A chief consequence of these models is that results and conclusions obtained through these models are probabilistic statements. This is, the conclusions are true only a specified percent of the time. Such models have been well accepted in several branches of scientific and management endeavors. Until recently, except for regression models, such models have not been significantly used in air pollution problems. Of late their use in air pollution problems has been increasing and such models appear to have a great potential for future work.

2) Deterministic Models—A deterministic model is one which does not have a probabilistic structure imposed on it. The models which fall under this category are
   a) Diffusion models
   b) Optimization models

The dispersion and the diffusion models have been well discussed in the literature, and are based upon idealizations of the physics of air transport and fluid flow. Most of the dispersion type models include meteorological variables and are useful as predictive models. Since these models involve a good deal of idealizations of the physical processes, a great deal of care should be exercised in making decisions based on such models.

Most of the optimization models used in air pollution problems are deterministic in nature. Functionally they fall under the category of models for control of air pollutant emissions, though they have also been used for problems involving the allocation of funds for air pollution control programs. The optimization models are of use to management for making broad based decisions and are extremely useful and practical. Despite these advantages, there seems to be a dearth of the use of such models in air pollution analysis and management problems. Once again, of late their use in air pollution problems has been increasing and it is likely that such models will be used more and more once their capabilities are discovered.

This section is concluded by emphasizing the following points.

1. For air pollution prediction, monitoring and control problems, mathematical models serve as valuable aids which have to be used judiciously.
2. The air pollution phenomenon is extremely complex; it is a func-

tion of the meteorology, climatology, the demography and the urbanization of a certain area. What is more, some of these variables interact with each other and are also time dependent. Mathematical models are idealizations or simplifications of reality and are developed under specific assumptions. In using these models one should ensure that the assumptions are adhered to and one should treat the results and conclusions from these models with extreme care.

E. Relationships Between Models and Decisions

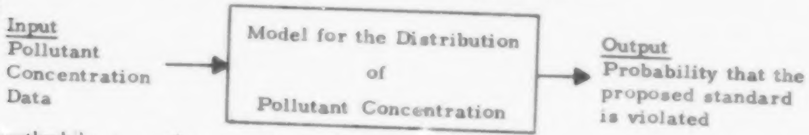
Before going into the specific details regarding the proposed models, it is worthwhile to point out how these models relate to each other and how they can aid a decision maker at the State or the Federal level. This is possible since the models have been functionally categorized as predictive and forecasting or control and monitoring. As far as a decision maker in the Federal or State government is concerned, the control and monitoring type models are of the greater interest to him. As will be shown later, the prediction models could serve as inputs to the monitoring and control models. A decision maker concerned with the setting of pollutant standards will be most concerned with predictive and forecasting models. Of course, as pointed out before, the use of these models for specific management type tasks is not clear cut. One may wish to use some or all of the above models based on his specific need. As a rule, these models which are designed for more general use can be adapted to one's situation depending on the problem at hand. In view of this it is important for a decision maker to understand the use, the capabilities and the limitations of each of the models discussed here.

We will next consider several illustrations, each addressing itself to a specific problem and then consider an illustration wherein several of these models could be used in the same problem.

1. A decision maker at the State or the Federal level is interested in evaluating or setting a standard for a certain pollutant. Specifically he wishes to know how often the observed concentration will violate the standard under question.

The decision maker will first collect some data on the pollutant under question and use the model for the distribution of the pollutant concentration. The following is a schematic of the steps, chart i.

i

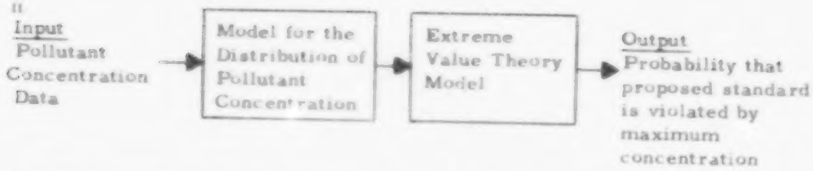


2. Suppose now, that this decision maker is satisfied with the standard discussed in 1 above. However, he wishes to test if the observed maximum concentration will violate this standard too often. As discussed before, air pollutant standards are often set for maximum concentrations observed for certain averaging times. Thus a question that may arise is, "how often does the observed maximum concentration for a specified averaging time exceed a given standard?"

The decision maker will now use in addition to the model for the distribution of pollutant concentrations an extreme value theory model. A schematic of the steps involved appears at the top of p. 74, chart ii.

3. It is conjectured that there are several controllable factors which contribute to the pollution of a certain geographic area. Some of these may be amount of industrial heating, the amount of traffic flow, number of industrial plants which emit effluents into the air, the quality of fuels used by these plants, the population of an area, etc. A decision maker wants to verify this conjecture, with the ultimate aim

ii


Input
Pollutant Concentration Data
Model for the Distribution of Pollutant Concentration
Extreme Value Theory Model
Output
Probability that proposed standard is violated by maximum concentration

that by controlling those factors that contribute the most to the pollution, he can improve the air quality. In effect the decision maker wishes to know which factors contribute to the air pollution and to what extent. He also wishes to know by what amount can the air pollution be reduced if he reduces the effect (the values) of these factors.

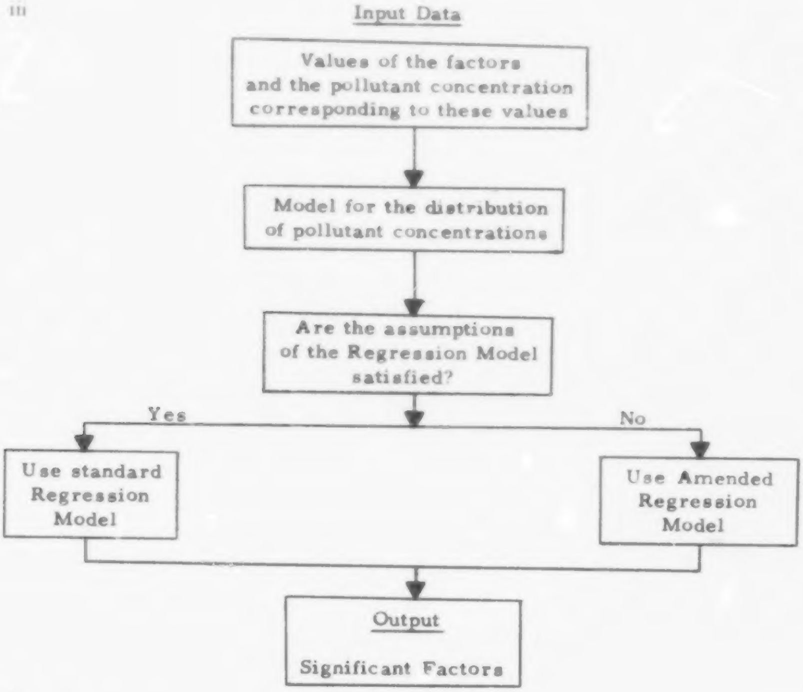
For this problem the decision maker could use a regression model, provided that the assumptions for regression are satisfied.

The following is a schematic of the steps, chart iii.

4. Now suppose that the decision maker discussed in 3 above would like to make broad predictions or forecasts of future pollution in his area based on the increase (or the decrease) of the values of his factors. For example, he may know that the population in his area is projected to grow by a certain amount or that due to the construction of a public transportation system in his area, there is going to be a decrease in the traffic. Based on these projections he may wish to predict the pollutant concentrations in his area.

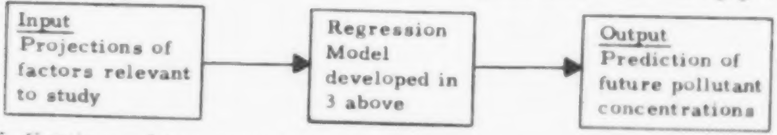
The decision maker could use the

iii


Input Data
Values of the factors and the pollutant concentration corresponding to these values
Model for the distribution of pollutant concentrations
Are the assumptions of the Regression Model satisfied?
Yes
No
Use standard Regression Model
Use Amended Regression Model
Output
Significant Factors

regression model discussed in 3 above to accomplish this, chart iv.

iv

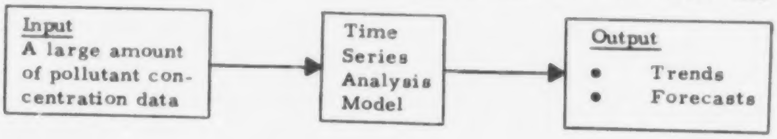

Input
Projections of factors relevant to study
Regression Model developed in 3 above
Output
Prediction of future pollutant concentrations

5. Certain regulatory measures have been instituted in a particular area, in an effort to lower the level of, say carbon monoxide concentrations. After a certain period has lapsed, that is, some time after these control measures have been introduced, a decision maker would like to ascertain if there is a downward trend in pollutant concentrations. He would also like to determine if this trend is going to continue into the future, assuming that there are no anticipated changes in the future which are going to increase the concentration levels.

An appropriate model that can be used here is a time series model. A time series model has fewer assumptions in it than a regression model. It is a model based on past data alone. it can detect trends, both cyclical or otherwise, and what is more, it can be used for forecasting and predicting. It is to be emphasized that a regression model is developed for the purposes of estimation and not for the purposes of predictions. However, with regression models one can perform predictions in a limited fashion. A time series model is designed purely for the purposes of forecasting and predicting.

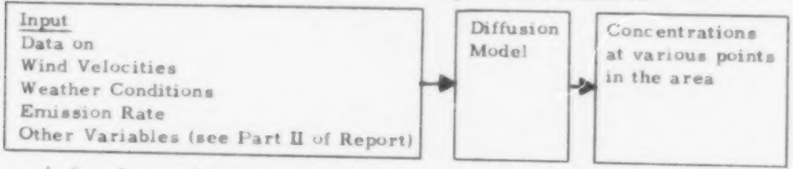
The following is a schematic of the steps involved, chart v.

v


Input
A large amount of pollutant concentration data
Time Series Analysis Model
Output
• Trends
• Forecasts

6. The city management council of a metropolitan area is considering giving permission to a power company to install a large generating plant in its vicinity. The generating plant will have a large stack (chimney) from which pollutants will be emitted into the atmosphere. When the wind velocity is low, these effluents could disperse within the metropolitan area increasing the amount of pollution in that area. The downtown part of the metropolitan area is densely populated and a city planner would like to know the ground level concentration due to effluents emitted by the stack The city planner has information about the wind velocities, weather conditions, etc. An appropriate model to use here is the Gaussian plume diffusion model, or other diffusion models, discussed in Part II.

The following is a schematic of the steps involved, chart vi.


Input
Data on
Wind Velocities
Weather Conditions
Emission Rate
Other Variables (see Part II of Report)
Diffusion Model
Concentrations at various points in the area

vi

7. A decision maker is faced with the problem of selecting air quality standards which minimize for the society the total cost of pollution. The total cost of pollution is the cost of pollution control. Problems of this nature involve a complex decision making process, a conclusion from which has several consequences. Problems of this nature involve several conflicting objectives and are not direct and

straightforward. Several factors, economic, social and political have to be considered in decisions of this type.

Models that can be used here are the optimization models or the decision theory models. The following chart abstracted from Drake, Keeney and Morse [3] depicts the flow of steps involved, chart vii.

The outputs from the chart below can be used as inputs to an optimization model.

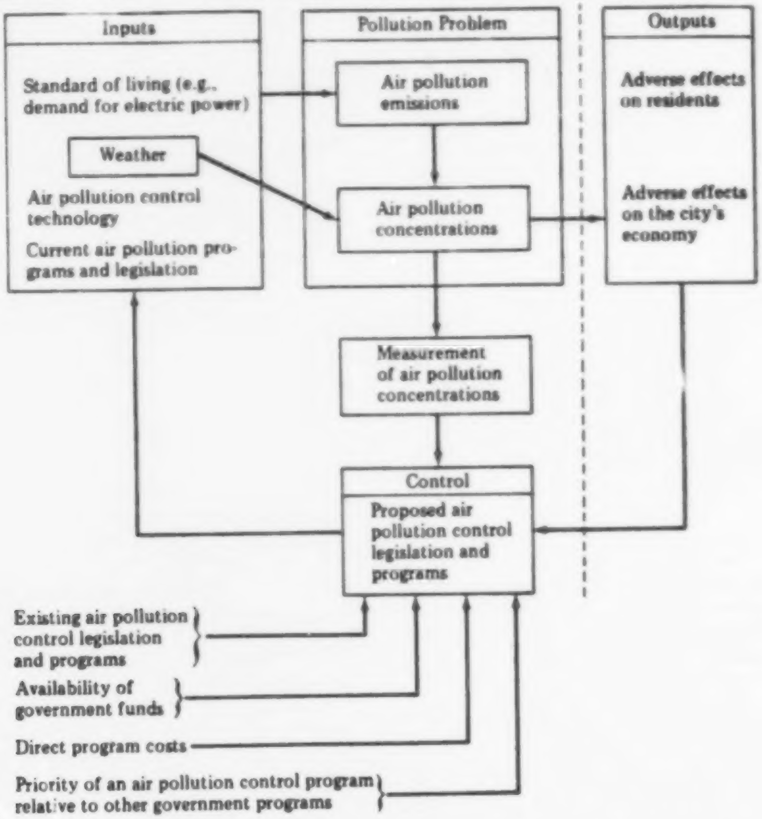
In any decision making environment, several of the models discussed here may have to be simultaneously employed as indicated below, chart viii. These models are mere aids to the general decision making process, and the outputs from these are not to be considered as all encompassing.

The arrows between the regression models, the time series analysis models, the extreme value theory models and the diffusion models indicate that they may have to interact with each other to obtain consistent results. The dotted boundary encloses the stochastic models.

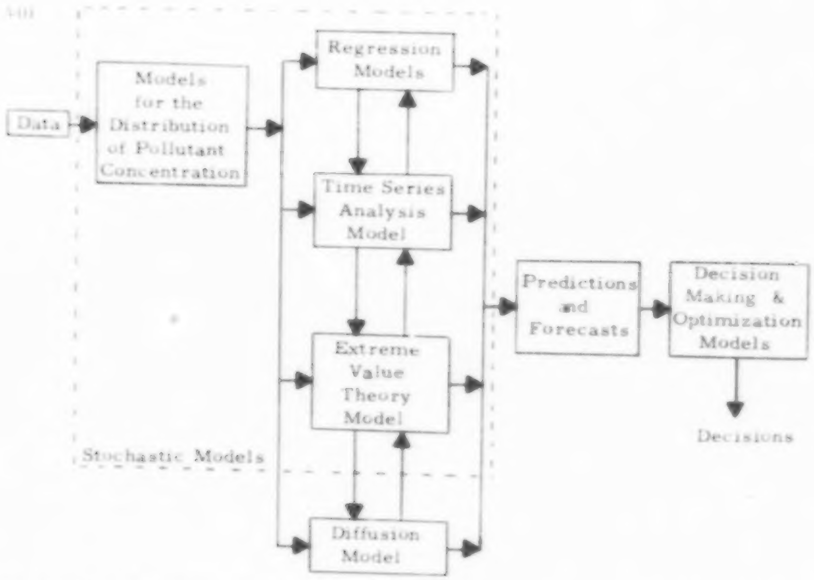
II. DISCUSSION OF SPECIFIC MODELS USED IN AIR POLLUTION

In this section we outline some specific stochastic and deterministic mod-

vii


Inputs
Standard of living (e.g., demand for electric power)
Weather
Air pollution control technology
Current air pollution programs and legislation
Pollution Problem
Air pollution emissions
Air pollution concentrations
Outputs
Adverse effects on residents
Adverse effects on the city's economy
Measurement of air pollution concentrations
Control
Proposed air pollution control legislation and programs
Existing air pollution control legislation and programs
Availability of government funds
Direct program costs
Priority of an air pollution control program relative to other government programs

General model for evaluating air pollution control programs.
Note: the arrow symbol ⟶ reads "influences."
Source: Drake, Keeney and Morse [3].


Data
Models for the Distribution of Pollutant Concentration
Regression Models
Time Series Analysis Model
Extreme Value Theory Model
Diffusion Model
Stochastic Models
Predictions and Forecasts
Decision Making & Optimization Models
Decisions

els which have been used in air pollution analysis. We shall point out the inputs and the outputs of these models, the major assumptions underlying these models and the nature of conclusions from such models. For convenience, we shall first discuss those models which have a probabilistic structure on it, and then discuss the deterministic models. How these models can be used has been discussed above. In the interest of exposing to the reader specific developments in mathematical models for air pollution, this section is detailed and technical.

A. Stochastic Models

We first introduce the following notations and conventions, which are necessary to discuss the mathematical models that are mentioned here.

Let the concentration of a certain pollutant, say sulfur-dioxide or carbon monoxide, etc., be denoted by X. From now on, X is a generic name given to the concentration of any pollutant. We shall assume that X varies according to some probabilistic law, one that we would like to know or determine. In probability theory X is known as a random variable.

Since X varies, it takes several values, and let $x_1, x_2, \ldots, x_n$ denote the various values it takes. For example, $x_1, x_2, \ldots, x_n$ may be the observations from the CAMP data discussed in I.B before. We shall denote the particular values which X takes by x; this is a convention adopted in the interest of generality and brevity. The probabilistic law according to which X varies can be described by the following statement:

$$\text{Probability}\ (X \leq x) = F(x).$$

F(x) is known as the *distribution function* of X, and it varies between 0 and 1. It can be shown that F(x) completely describes the probabilistic behavior of X.

If X can take all values in an interval (finite or infinite), X is said to be a continuous random variable. We will assume, that as far as air pollution problems are concerned, X is always a continuous random variable.

If X is a continuous random variable,

it can be proved that there exists a function f(x) such that

$$F(x) = \int_{-\infty}^{x} f(s)\, ds.$$

f(s) is known as the *probability density function* of X and it can be used to describe the probabilistic behavior of X.

In order to be able to summarize the behavior of X, a varying quantity, we introduce the notion of the mean and the variance.

The mean of X, to be denoted by μ is defined as

$$\mu = \int_{-\infty}^{\infty} x f(x)\, dx.$$

μ is a measure of its central tendency.

The variance of X, denoted as σ^2 is defined as

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx.$$

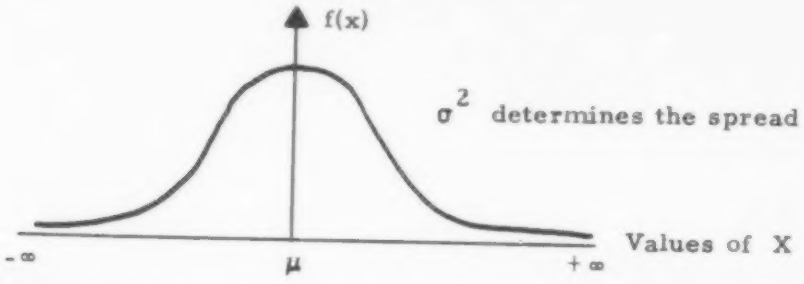
σ^2 is a measure of the spread of X.

For example, if

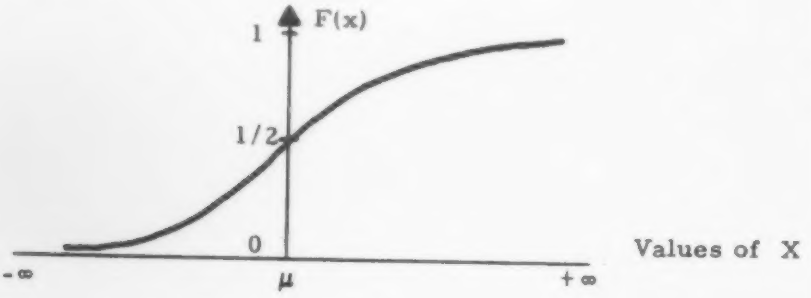
$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty$$

then, f(x) is said to have the well known normal or the bell-shaped distribution.

For illustrative purposes, we sketch below the probability density function, and the distribution function of a normal distribution.



The Probability Density Function of X



The Distribution Function of X

I. *Models for the Distribution of Pollutant Concentrations*

Models for the distribution of pollutant concentrations are of fundamental interest. It is these models which enable us to introduce a specific probabilistic structure for the stochastic models introduced later. The determination of an appropriate model for the distribution of a pollutant concentration should be construed as the first step in any statistical analysis and decision-making procedure. Such models are very useful for the setting of air quality standards. For example, Larsen's [1] work on the lognormal distribution falls in this category of models.

The determination of the distribution of a pollutant concentration is essentially determining the functional form of f(x) or F(x) discussed before. Thus, we are interested in determining the probabilistic law which explains the variation of the concentration of a particular pollutant.

In order to be able to do this, we will have to make several assumptions.

As assumed before, let $x_1, x_2, x_3, \ldots, x_n$ be the values of the concentration for a particular pollutant for some averaging time, say t_o.

Assumptions

i) We shall assume that the values $x_1, x_2, \ldots, x_n$ are *statistically independent;* that is, they are not correlated.

ii) We shall assume that the n observations $x_1, x_2, \ldots, x_n$ are taken from some unknown distribution F(x) which does not change in time.

Discussion of the Assumptions

Recall, that in Section I.B we had denoted the successive values of the pollutant concentrations from the **CAMP** data by $x_{t_1}, x_{t_2}, \ldots, x_{t_n}$, where $t_1 < t_2 < t_3 \ldots < t_n$. For this data the independence assumption does not apply, and we shall treat this kind of data as a time series, which we will discuss later.

The point to be made here is that for small averaging times, say 5 minutes or so, due to the effects of the weather, the assumption of the independence of the x_i's may not be true. However, for large averaging times, the assumption of independence is more reasonable and we restrict ourselves to averaging times of say eight hours or more. Thus for our purposes, the $x_1, x_2, \ldots, x_n$ denote concentrations for averaging times of 8 hours or more, and can be assumed to be independent.

The assumption that F(x) does not change in time is also a questionable one. Clearly, for large averaging times, due to the cumulative effects of the air pollutants which are being discharged into the atmosphere and the other weather variables, it is possible that the mean and the variance of F(x) may change. However, there are several empirical studies by Larsen [1] which verify that the distribution does not change for averaging times which are not too large, say one week or less.

Clearly, the above two assumptions present the analyst with conflicting choices regarding the selection of a suitable averaging time for analyzing his data. For the kind of analysis suggested here, we suggest that the analyst choose an averaging time which is not too small (less than 8 hours) to violate (i) and not too large (greater than one week) to violate (ii).

Several models for the distribution of pollutant concentrations have been suggested in the air pollution literature. The most prominent one of these seems to be the lognormal distribution advocated mainly by Larsen and his colleagues [1]. Since air pollution data is highly skewed, one might easily conjecture that the data might as easily be fit with a gamma, a Weibull or any other skewed distribution. The question as to what distribution of pollutant concentrations is a reasonable one is a highly debated item.

There are several procedures which can be used to verify if the observed data on pollutant concentrations can be reasonably well explained by a hypothesized probability distribution function. These procedures are well described in any text book on applied statistics. The procedures that are most commonly employed are:

i) Probability plotting procedures.
ii) Chi-squared goodness of fit test procedures.
iii) Kolmogorov-Smirnov tests.

From a practical point of view, probability plotting procedures seem to be the most expedient and simple. In this procedure, either the observed data, or simple transformations on it, such as logarithms, are plotted against estimates of the distribution function F(x) on special paper, known as probability paper. For $X = x_i$, a reasonable (unbiased) estimator of F(x) is $\frac{i}{n+1}$, where the x_i refers to the i^{th} smallest observation. Probability paper for commonly used distributions is commercially available, and the acceptance of a hypothesized distribution is usually based on the linearity of the aforementioned plot.

It is to be emphasized here that the acceptance of a hypothesized distribution based on probability plotting procedures is not to be treated as all conclusive and final.

There is a lot of judgment involved in entertaining a decision of the type discussed here; namely, accepting a hypothesized distribution as a reasonable one to use. Firstly, the question of concluding linearity may be a subjective one despite the fact that there are statistical procedures for assisting this procedure. Secondly, the observed data may be incorrectly observed, i.e., it may not be reliable, or it may not be available in a sufficiently large volume for an analysis of this type. It is also possible that the same set of data may be reasonably well explained by two or more conjectured (hypothesized) distributions. This is because the differences between some of the well known distributions are dominant at the tails of these distributions and the available data may not cover the entire range of the distribution. In such cases what may be needed is more data which are spaced to cover a large range of the distribution, together with technical opinions of specialists who may be able to conjecture the appropriateness of a distribution based on the physical processes which generate it.

Once a distribution for pollutant concentrations has been adopted, it must be borne in mind that this is merely an idealization of the distribution which may be quite complicated.

a) *The Lognormal Distribution*

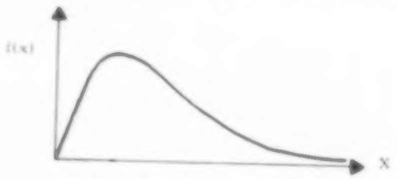
As mentioned before, this distribution has been widely used in an analysis of air pollution data. Of course, there is much debate as to whether this distribution is a reasonable one for all air pollutants.

A continuous random variable X is said to have a lognormal distribution if

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma x}\, e^{-\frac{1}{2}\left(\frac{\log_e x - \mu}{\sigma}\right)^2}, \quad 0 \leq x < \infty$$

It is easy to verify that if $Y = \log_e X$, then Y has a normal distribution with a mean μ and a variance σ^2.

The probability density function of a lognormal distribution is shown below.



It can be verified that the mean of this distribution is

$$e^{\mu + \sigma^2/2}$$

and that its variance is

$$e^{2\mu+\sigma^2}(e^{\sigma^2} - 1).$$

From a practical point of view, often in the analysis of air pollution data, one needs to estimate the mean pollutant concentration and the variance of the pollutant concentration for some pollutant which is conjectured to have a lognormal distribution.

If $x_1, x_2, \ldots, x_n$ denotes the air pollutant concentration values discussed before, then an estimate of the mean concentration is given as

$$\hat{\mu} = \sum_{i=1}^{n} \log x_i / n,$$

and an estimate of the variance of the concentrations is

$$\sigma^2 = \sum_{i=1}^{n} (\log x_i - \hat{\mu})^2 / (n - 1).$$

The next question is to determine if the data $x_1, x_2, \ldots, x_n$ does have a lognormal distribution. Probability plots for the lognormal distribution can be obtained by plotting the logarithms of the observed data, namely, $\log_e x_1, \log_e x_2, \ldots, \log_e x_n$ on normal probability paper. Such a plot for some particulate data is shown in the accompanying figure (Figure 1). Since the points are not quite on a straight line, it is difficult to conclude if this particulate data can be described by a lognormal distribution. The lognormal distribution is discussed in good detail by Aitchison and Brown [4]. Larsen [1] has concluded that concentrations are approximately lognormally distributed for all pollutants, in all cities for all averaging times. Based on this conclusion several standards for relating air quality to averaging times have been adopted.

b) *The Gamma Distribution*

This distribution for pollutant concentrations has been conjectured by

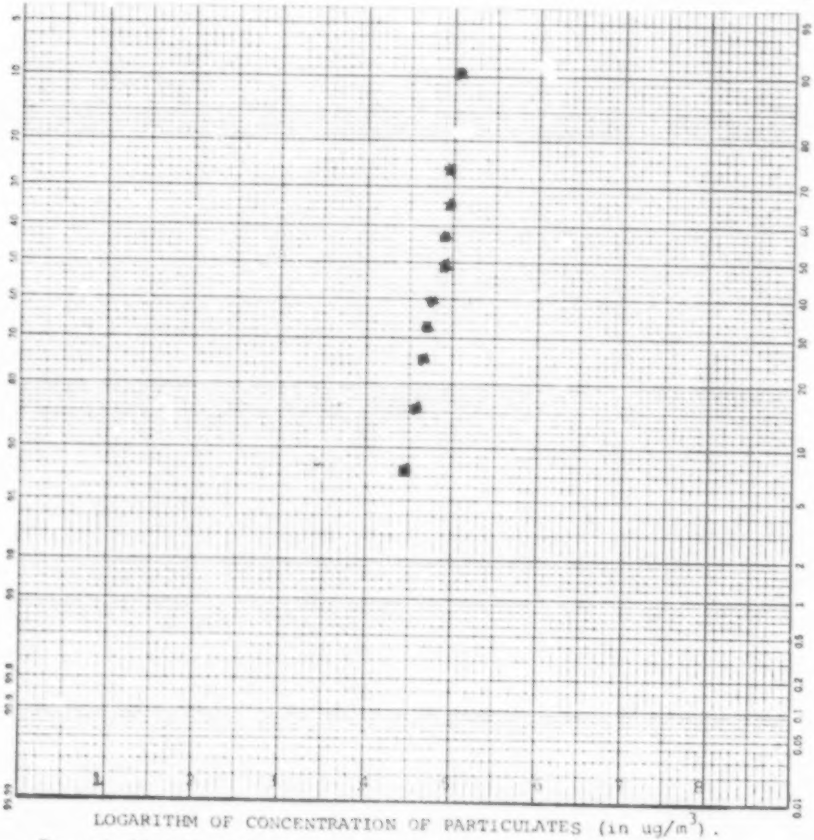
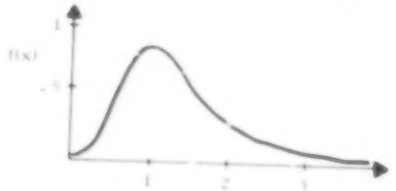


FIGURE 1—*Plot of logarithm of particulate data (in μg/m³) on normal probability paper*

Barlow and Singpurwalla [5]. This conjecture is based on the observation that air pollution data being highly skewed, any skewed distribution may be a reasonable one to consider.

The probability density function for the gamma distribution is shown below.

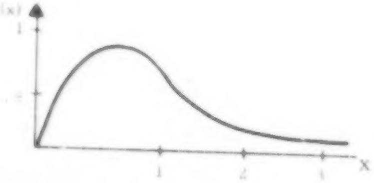


Methods for estimating the parameters of the gamma distribution are given by Choi and Wette [6]. Probability plots for the gamma distribution can be obtained by using the procedure suggested by Wilk, Gnanadesikan and Huyett [7]. The procedure is rather involved to be discussed here.

c) *The Weibull Distribution*

This distribution for pollutant concentrations has been conjectured by Barlow [8]. Once again, this conjecture is based on the skewness of the observed data.

The plot of a typical Weibull distribution is shown below.



Estimators of the parameters β and δ of this distribution are obtained by using tables provided by Nancy Mann (1967).

Probability plots for the Weibull distribution can be obtained by plotting

$$\log\left[\log\left\{\frac{1}{1-\left(\frac{i}{n+1}\right)}\right\}\right] \text{ versus}$$

$\log x_1 \leq \log x_2 \leq \ldots \ldots \leq \log x_n$, where $x_1 \leq x_2 \leq x_3, \ldots \ldots \leq x_n$ are the ordered observed pollutant concentrations. Such a plot for the particulate data discussed before is shown on the accompanying figure (Figure 2).

2. *Regression Models in Air Pollution Analysis*

Because of the general applicability of regression models to a wide variety of situations, these models have been often used in air pollution analysis. For example, regression models have been used to evaluate and forecast expenditures due to air pollution abatement strategies, to discuss the effects of the various variables on air pollution, and a host of other applications. More recently, regression models are used to construct what are known as "smog diagrams" in an analysis of air pollution data. These will be discussed later. Regression models have also been used to evaluate certain health effects due to pollutants in the atmosphere. It is to be stated that regression models can be used as diagnostic tools or as predictive tools in air pollution analysis.

In view of the several applications of regression models to problems in air pollution we shall first review some assumptions in regression models and establish some convention and notations common to such models.

a) *Conventions and Notation*

Let $Y_1, Y_2, \ldots, Y_k$ be known variables, X an observable random variable and $\beta_0, \beta_1, \ldots, \beta_k$ unknown parameters. Then, the relationship

$$X = \beta_0 + \beta_1 Y_1 + \beta_2 Y_2 + \ldots + \beta_k Y_k + e,$$

is defined as the general linear model.

For example, X could be the observable concentrations of a certain pollutant, and $Y_1, Y_2, \ldots, Y_k$ may be observable values of the variables such as wind speed, average temperature, wind direction (in radians), average pressure, etc. which influence the concentrations of a given pollutant.

Our objective is to obtain estimates of the parameters $\beta_0, \beta_1, \ldots, \beta_k$, so that given some values of $Y_1, Y_2, \ldots, Y_k$, the value of X can be predicted. Of course, a knowledge of $\beta_0, \beta_1, \ldots, \beta_k$ will also help us to assess the contribution of the variable Y_i to X.

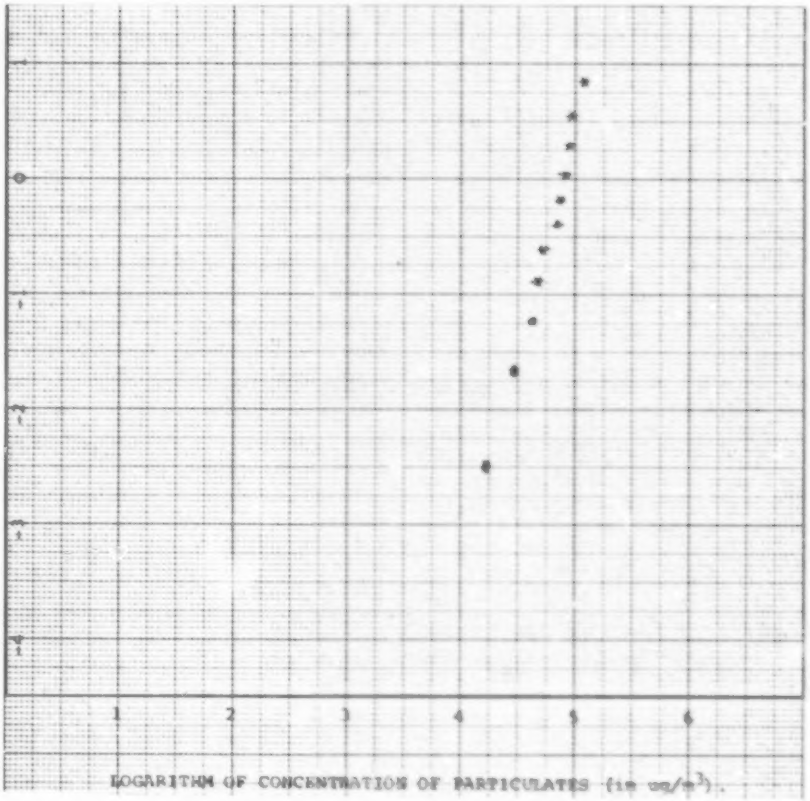


FIGURE 2—*Weibull plot of particulate data (in μg/m³) on linear by linear scale*

Let Y_{ij}, i = 1, 2, . . ., n denote the different values which variable Y_j, j = 1, 2, . . ., k takes. When the variables Y_j, j = 1, 2, . . ., k take values i = 1, 2, . . ., n, let X take a value X_i.

That is, when we set $Y_1 = Y_{1i}$, $Y_2 = Y_{2i}$, . . ., $Y_k = Y_{ki}$, we observe $X = X_i$.

Our data leads us to a following representation of the model

$$X_1 = \beta_0 + \beta_1 Y_{11} + \beta_2 Y_{21} + \ldots + \beta_k Y_{k1} + \epsilon_1$$

$$X_2 = \beta_0 + \beta_1 Y_{12} + \beta_2 Y_{22} + \ldots + \beta_k Y_{k2} + \epsilon_2$$

$$\vdots$$

$$X_n = \beta_0 + \beta_1 Y_{1n} + \beta_2 Y_{2n} + \ldots + \beta_k Y_{kn} + \epsilon_k.$$

For example, the first equation states that when variable Y_1 takes a value Y_{11}, and variable Y_2 takes a value Y_{21}, . . ., and variable Y_k takes a value Y_{k1}, the random variable X takes a value X_1.

In regression theory, the X is referred to as the *dependent* variable and the Y_j's, j = 1, 2, . . ., k as the *independent* variables. In the illustration above we have n sets of data, and it is necessary that $n \geq k + 1$.

b) *Principle of Least Squares*

The principle of least squares, or regression requires us to pick those values of the β_i's, namely $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ such that

$$\sum_{i=1}^{n} (X_i - \hat{X}_i)^2$$

is minimized, where

$$\hat{X}_i = \hat{\beta}_0 + \hat{\beta}_1 Y_{1i} + \hat{\beta}_2 Y_{2i} + \ldots + \hat{\beta}_k Y_{ki}, \quad i = 1, 2, \ldots, n.$$

The above minimization procedure is quite straightforward and can be routinely performed on a computer.

c) *Variations in the Basic Model*

The model presented in section (b) is known as a linear model because it is linear in the parameters $\beta_0, \beta_1, \ldots, \beta_k$.

In view of this, any non-linearity in the Y_j, such as $\log Y_j$, or $\sqrt{Y_j}$, or sine Y_j, for any of all j would still retain the basic character of the model as long as there are no parameters which are non-linear. For example, if it were conjectured the sulfur dioxide concentrations X_i are linear functions of the logarithm of the wind speed $\tilde{Y}_1$ and the square root of the average temperature $\tilde{Y}_2$, then the model

$$X_i = \beta_0 + \beta_1 \log_e \tilde{Y}_{1i} + \beta_2 \sqrt{\tilde{Y}_{2i}}, \quad i = 1, 2, \ldots, n$$

would be still a linear model, since a transformation $Y_1 = \log_e \tilde{Y}_1$ and $Y_2 = \sqrt{\tilde{Y}_2}$ would give us a linear model

$$X_i = \beta_0 + \beta_1 Y_{1i} + \beta_2 Y_{2i}, \quad i = 1, 2, \ldots, n.$$

In several instances, a non-linear model can be cast into the framework of a linear model by simple transformations. For example, suppose that based on some diffusion equations it is conjectured that the carbon monoxide concentration X is related to the average temperature as Y_1

$$X_i = e^{-(\beta_0 + \beta_1 Y_{1i})}.$$

A logarithmic transformation leads us to

$$-\log X_i = \beta_0 + \beta_1 Y_{1i},$$

again a linear model.

d) *Step-wise Regression Procedure*

The step-wise regression procedure is so commonly used in practice, including problems pertaining to air pollution that it deserves special mention.

In practice, one rarely knows the form of the model he should use. He may consider a certain number of variables to be the pertinent ones but he may not be sure if all the variables that he has considered are really relevant or not. The model discussed in 2(a) is one that we may be willing to tentatively entertain, or one which because of reasoning based upon the physical nature of the problem at hand, we are willing to adopt.

The step-wise regression procedure is a procedure which allows us to systematically pick a variable from a given number of variables, and include it in the model or exclude it from the model based upon the contribution of that variable.

The step-wise regression procedure first picks that variable whose contribution to the model is the largest and then picks another variable from the remaining ones to see if its contribution is significant. A nice feature of the step-wise procedure is that it eliminates from the model any variable which it had previously included in the model if the contribution of this variable is minimized by the choice of subsequent models.

Computer programs routinely perform the step-wise regression procedure in a highly mechanized manner.

e) *Prediction Intervals from Regression Models*

A $(1 - \alpha)\%$ prediction interval can be obtained for a new predicted value of X, say x_ℓ for values of Y_j, at $Y_{j\ell}$, $j = 1, 2, \ldots, K$, and $\ell \neq i$.

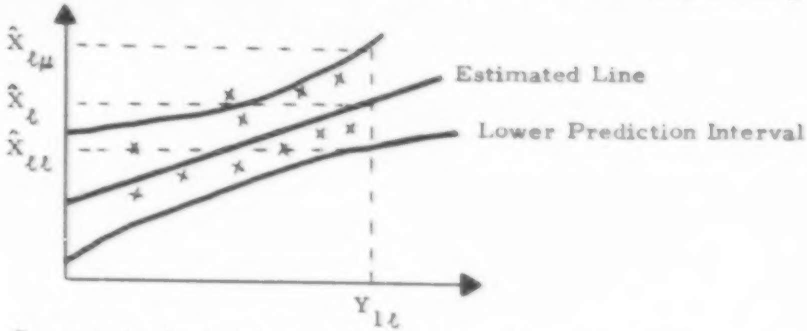
We illustrate by considering a simple linear regression model in two variables

$$X_i = \beta_0 + \beta_1 Y_{1i} + \epsilon_i, \quad i = 1, 2, \ldots, n.$$

In the figure below, the data is indicated by crosses, the estimated line based on least squares

$$\hat{X}_i = \hat{\beta}_0 + \hat{\beta}_1 Y_{1i}$$

is also shown, and the hyperbolic curves represent the $(1 - \alpha)\%$ prediction intervals. It can be verified that these intervals are the narrowest within the range of the data, especially at the center and spread out as we move away towards the ends. It is for this reason that it is not advisable to make predictions from regression lines for points too far away from the range of the data. The hyperbolic nature of the prediction intervals is due to the assumption of normality of the error.



For example, if we wish to predict the value of X at $Y_i = Y_{1\ell}$, where $Y_{1\ell} \neq Y_{1i}$, $i = 1, 2, \ldots, n$, then $\hat{X}_\ell$ is the predicted value. $\hat{X}_{\ell\ell}$ is the lower prediction limit and $\hat{X}_{\ell u}$ is the upper prediction limit.

The interpretation of the $(1-\alpha)\%$ prediction limits is that the probability that the random interval $(\hat{X}_{\ell u} - \hat{X}_{\ell\ell})$ contains the true value X_ℓ is $(1-\alpha)\%$.

f) *Applications of Regression Analysis for Constructing Smog Diagrams*

As mentioned before, regression analysis has been widely used in air pollution problems. We illustrate here an application of this technique to construct what are kown as "smog diagrams." A smog diagram is a pictorial depiction of the mathematical relationship between oxidants, hydrocarbons and nitrogen oxides. These pollutants are responsible for the smog in a particular area.

The results reported here are based on a study conducted by Merz et al. [9] for the downtown Los Angeles air monitoring station. Data for the months of August, September and October for the years (1962–1969) were used. The 6–9 A.M. three hour averages were used for total hydrocarbon and nitrous oxide concentrations. The maximum hourly average concentration occurring on that day at the downtown Los Angeles station was used for the total oxidant.

Using a step-wise regression procedure the logarithm of the oxidants were fitted by a quadratic model in hydrocarbons (HC) and nitrogen oxide (NO_x).

The model is written as

$$X = \log(\text{oxidant}) = \beta_0 + \sum_{i=1}^{8} \beta_i Y_i$$

where $Y_1 = \log(NO_x/17.5)$, $Y_2 = \log(HC/4.6)$, $Y_3 = Y_1^2$, $Y_4 = Y_1 \, Y_2$, $Y_5 = Y_2^2$, $Y_6 = Y_1^2$, $Y_7 = Y_1 \, Y_2^2$ and $Y_8 = Y_1^2 \, Y_2^2$; 17.5 and 4.6 are constants used for computational advantages.

Results of the Regression

The step-wise regression procedure eliminated all the variables except Y_1 and Y_2 to yield the following estimated relationship:

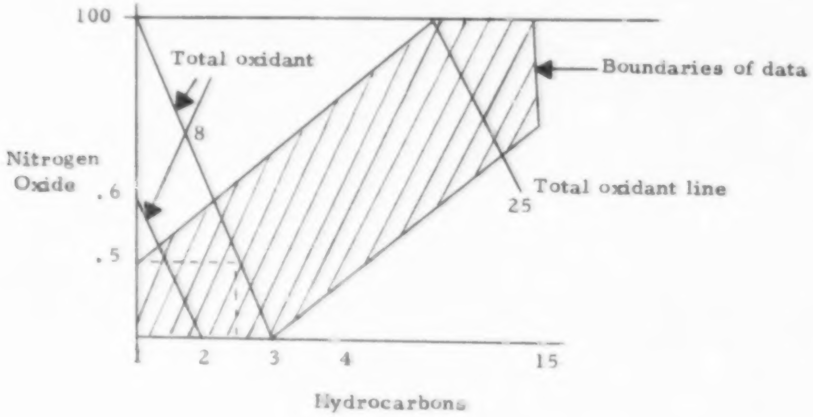
$$\text{oxidant} = e^{2.6}\left(\frac{HC}{4.6}\right)^{.54}\left(\frac{NO_x}{17.5}\right)^{.15}$$

or

$$\log(\text{oxidant}) = 2.6 + .54\log\left(\frac{HC}{4.6}\right) + .15\log\left(\frac{NO_x}{17.5}\right).$$

Thus, estimators of the parameters β_0, β_1 and β_2 are 2.6, .54 and .15 respectively.

The following diagram depicts this relationship; this diagram is called the smog diagram.



“Smog Diagram”

As indicated by the dotted lines, when the level of hydrocarbons is at 2.5 and the nitrogen oxide is at .5, the total oxidant takes a value of 8. The shaded area represents the region covered by the data, and predictions within this region are the most reliable.

3. *Models for Predicting Maximum Concentrations*

In Section I.B we had mentioned that the maximum pollutant concentration $\eta_{K,n}$ of a particular pollutant is of interest. We had also stated several State and Federal standards for air pollutant concentration for a specified averaging time.

In an attempt to set standards, based on maximum concentrations, which are reasonable, or in an attempt to investigate the reasonableness of a specified standard, we resort to the use of extreme value theory.

The application of extreme value theory to air pollution problems is relatively new. It was first advocated by Singpurwalla [10] and then followed up by Barlow [8]. Barlow and Singpurwalla [5] have also indicated how the extreme value theory can be used in problems where the data may be correlated, a situation typical with air pollution data. These and other details are discussed in the following sections.

a) *A Statistical Model for Maximum Concentrations*

Reverting to the notation established in Section I.B, let X_{t1}, X_{t2}, . . . , X_{tn} be the values of the concentration for a particular pollutant, for a five minute averaging time.

Let us assume for now, the following:

i) the values X_{t1}, X_{t2}, . . . , X_{tn} are statistically independent (i.e. not correlated),
ii) the values X_{t1}, ${}_{t2}$, ... , X_{tn} are taken from a fixed distribution F(x) which does not change.
iii) the distribution F(x) is either a lognormal, a Weibull, a gamma, a normal or an exponential distribution.

Discussion of the Assumptions

As we have pointed out in Section II.A.1, the assumptions (i) and (ii) may be difficult to realize in all the cases. Despite this difficulty, we continue with the discussion here, because we can prove that the results obtained by assuming independence serve as reasonable bounds for similar results under the non-independence case (Barlow and Singpurwalla [5]).

Assumption (iii) is very general and there should be no difficulty in realizing this assumption in practice. Of course, several other distributions can be included under assumption (iii) as long as these distributions belong to the domain of attraction of the limiting law discussed later in this section. The concept of the "domain of attraction" of a limiting law is quite mathematical and is omitted here; for more details on this we refer the reader to Gnedenko [11].

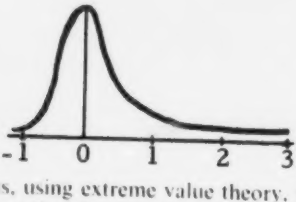
As discussed before, we next form averages of length k, and obtain $\eta_{k,n}$ as

$$\eta_{k,n} = \text{Max}\,\{k^{-1}(X_{t_1} + \ldots + X_{t_k}), \ldots, k^{-1}(X_{t_{n-k+1}} + \ldots + X_{t_n})\}.$$

Since $\eta_{k,n}$ is a function of the random variable X whose values are X_{t_1}, X_{t_2}, . . ., etc., $\eta_{k,n}$ is also a random variable whose probabilistic behavior is described by what is known as an extreme value distribution, if assumptions (i) and (ii) hold. If assumption (iii) holds, then as n gets large, and if k is moderate, i.e. if the averaging time is moderate

$$P\{\eta_{k,n} \le x\} = \exp\left[-\exp\left(-\left(\frac{x - \beta_{k,n}}{\alpha_{k,n}}\right)\right)\right] \quad -\infty < x < \infty,$$

where $\beta_{k,n}$ and $\alpha_{k,n}$ are parameters or constants, known as normalizing constants. The above result follows from the extreme value theory (Fisher and Tippett (1928)). We illustrate below a plot of the extreme value distribution



Thus, using extreme value theory, we can describe the random behavior of the maximum concentration $\eta_{k,n}$ as a function of the data, n, and the averaging time 5k. We illustrate an application of this in the following problem.

b) *Applications to Air Quality Data*

Studies indicate that plants grown in a particular area could be visibly damaged if the 8-hour average concentration of oxidant exceeded .03 ppm. Thus, if the air management board desired to prevent such damage they might set the air quality standard at a maximum value of 0.03 ppm for an 8-hour averaging time.

The question of how frequently the above standard will be violated and other related questions can be answered by determining the probability that the maximum concentration $\eta_{k,n}$ exceeds some value λ (say .03 ppm) for some averaging time t_a (say 8 hours). The constants $\beta_{k,n}$ and $\alpha_{k,n}$ are to be obtained from air pollution data, and are a function of the distribution function F(x) discussed in assumption (iii). For more details on this, the reader is referred to Singpurwalla [10], and Barlow and Singpurwalla [5]. Bounds on $\beta_{k,n}$ have been obtained by Barlow [8].

c) *Extensions for the Case of Dependent Data*

It was assumed in Section 3.a that the values $X_{t_1}, X_{t_2}, \ldots, X_{t_n}$ are independent. It was also mentioned that the assumption of independence gave us meaningful bounds for the distribution function of the maximum under the non-independence case. These results are highlighted in this section.

We shall now assume that the values

$$X_{t_1}, X_{t_2}, \ldots, X_{t_n}$$

are dependent (or correlated) such that the "autocorrelation function" is always positive and it damps out. The notion of an autocorrelation function is discussed in Section 4.e to be presented later.

The results that are given here are true under a much stronger notion of dependence known as "association." The concept of association is due to Esary, Proschan and Walkup [12] and is quite mathematical to outline here. However, it is true that a necessary condition for association is that the autocorrelation function be positive.

Barlow and Singpurwalla [5] prove that if the data $X_{t_1}, X_{t_2}, \ldots, X_{t_n}$ is associated, then

$$P\{\eta_{k,n} \le x\} \le \exp\left[-\exp\left(-\left(\frac{x-\beta_{k,n}}{\alpha_{k,n}}\right)\right)\right].$$

It is easy to recognize the fact that the right hand side of the above equation is the distribution function of the maximum concentration under the assumption of independence.

In words, the above result implies that when the air pollution data is correlated (which is often the case), then the probability that the maximum concentration of a specified pollutant for some averaging time, say 5k, is less than or equal to X, is *greater* than the number obtained through the right hand side expression.

Thus, the assumption of independence in extreme value theory, when the data are associated (positively correlated), gives us conservative results.

4. *Time Series Analysis Models*

Air pollution data being highly dependent (correlated), time series models for analyzing these data appear to be the most appropriate. Whereas regression models are used for diagnostic and perhaps for prediction purposes, time series models are mainly used for prediction purposes.

Time series models have been used in air pollution analysis to predict either future pollutant concentrations or trends for various pollutant concentrations in several cities. Time series models require a healthy amount of previous data but no model assumptions, such as linear models, quadratic models, etc. used in regression models. We shall outline in this section the various time series models that can be used in an analysis of air pollution data and indicate a specific example of its use.

a) *Statement of the Problem*

Let $Z_t, Z_{t-1}, Z_{t-2}, \ldots$, be a set of observations at equispaced time points $t, t-1, t-2, \ldots$. $Z_t, Z_{t-1}, Z_{t-2}, \ldots$, etc. is referred to as a time series. We need to predict $Z_{t+\ell}$, where t denotes the present time, and $\ell = 1, 2, \ldots$, denote the future points in time. ℓ is known as the "lead time" for which the forecast is required, and if $\hat{Z}_t(\ell)$ denotes the forecast of $Z_{t+\ell}$, the $\hat{Z}_t(\ell)$ is known as the forecast at origin t for lead time ℓ.

For example, $Z_t, Z_{t-1}, Z_{t-2}, \ldots$, could represent the concentrations of a particular pollutant at times $t, t-1, t-2, \ldots$ etc., where the t's could be spaced 5 minutes apart. Thus the CAMP data could be considered to be a time series.

Loosely speaking a "stationary time series" is one which is in equilibrium about a constant mean, say μ. A "non-stationary" time series is one which is not stationary. Typical examples of non-stationary time series are stock prices, uncontrolled processes and even pollutant concentrations where no control policies exist.

We assume that the observations $Z_t, Z_{t-1}, Z_{t-2}, \ldots$, which are highly dependent are generated by a series of shocks $A_t, A_{t-1}, \ldots$. We further assume that the $A_t, A_{t-1}, \ldots$ etc. are independent random observations from a distribution which is normal with a mean zero and a variance σ_2. The $A_t, A_{t-1}, \ldots$, values are also known as the "white noise" process.

This concept is perfectly general, and difficult to dispute. Thus as far as the assumptions in time series are concerned, we need not worry much about their violation, since they are quite general. The assumptions stated here are more of a concept, an idealization, and are thus not restrictive.

b) *Some Stationary Time Series Models*

i) The Autoregressive Model

These models are very useful in practice, including air pollution problems.

Let us denote $Z_t - \mu$ by $\tilde{Z}_t$, $Z_{t-1} - \mu$ by $\tilde{Z}_{t-1}$ and so on. What enables us to do this is the stationary property which assumes the existence of a fixed mean μ.

Then, if $\phi_1, \phi_2, \ldots, \phi_p$ are any constants the model

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + \ldots + \phi_p \tilde{Z}_{t-p} + A_t$$

is known as an autoregressive process of order p. In this process, we idealize that the current value of the process Z_t is the sum of the p previous values of the process, each multiplied by a parameter ϕ_i, $i = 1, 2, \ldots p$.

The first and the second order autoregressive processes are most commonly used in practice.

ii) The Moving Average Model

The process

$$Z_t = A_t = \theta_1 A_{t-1} - \theta_2 A_{t-2} - \ldots - \theta_q A_{t-q}$$

is known as the moving average model of order q.

iii) The Mixed Autoregressive-Moving Average Model

In order to obtain parsimony, that is, in order to explain a large class of models with as few a number of parameters as is possible we introduce what is known as the autoregressive, moving-average process, abbreviated as the ARMA(p,q) process. In this process we have

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \ldots + \phi_p Z_{t-p} + A_t - \theta_1 A_{t-1} - \theta_2 A_{t-2} - \ldots - \theta_q A_{q-t}.$$

In practice, both p and q are rarely greater than two.

c) *Non-Stationary Time Series Models*

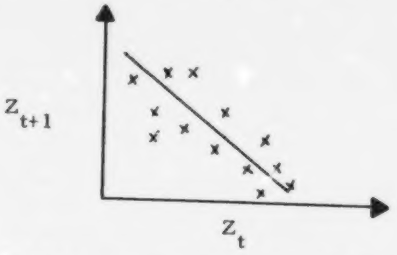
To introduce non-stationary considerations in the model we consider the d^{th} difference to the Z_t values. That is if we take the d^{th} differences of our data, we obtain an ARMA(p,q) model. If $d = 0$, the process reduces to an ARMA process.

The ARMA process is a very general process which includes an ARMA(p,q) process when $d = 0$, an AR(p) process when p and d are zero, and an MA(q) process when p and d are zero.

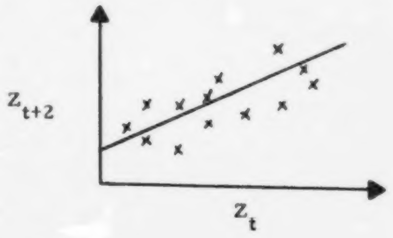
d) *The Notion of An Autocorrelation Function*

In order to identify a time series model, the autocorrelation function is used. Consider the pairs of values (Z_t, Z_{t+1}) and (Z_t, Z_{t+2}) for all values of t.

If we plot the pair (Z_t, Z_{t+1}) for all values of t and if we observe the following plot, we conclude that the pair (Z_t, Z_{t+1}) is negatively correlated.



Now suppose that we plot the pair (Z_t, Z_{t+2}) for all values of t, and if we observe the following plot, we conclude that the pair (Z_t, Z_{t+2}) is positively correlated.



The covariance between Z_t and Z_{t+k} is known as the "autocovariance" at lag k. It is denoted by γ_k. When $k = 0$ the autocovariance reduces to the variance.

The "autocorrelation" at lag k is the ratio of the autocovariance at lag k to the variance; it is denoted by ρ_k. Thus

$$\rho_k = \frac{\gamma_k}{\gamma_0}.$$

A plot of ρ_k versus k, for $k = 0, 1, 2, \ldots$ is known as the "autocorrelation function." Clearly, the autocorrelation function starts at plus or minus one and tails off as k increases. The autocorrelation function gives us an idea as to how the observations are correlated with each other. That is, are successive values dependent on each other or not, and if they are dependent, the nature of this dependence is important.

There are several practical considerations to be taken into account for estimating ρ_k. We outline these below:

i) n should be at least 50
ii) K should not be larger than n/4
iii) it is enough to round off the estimates of the autocorrelation function to two decimals

In the accompanying figures we present plots of two time series and their autocorrelation functions. These plots have been abstracted from Barlow and Singpurwalla [5].

e) *Forecasting Using Time Series*

It was stated before that time series models are mainly used for prediction purposes. Specifically, we had stated that given pollutant concentrations,

$$Z_t, Z_{t-1}, Z_{t-2}, \ldots$$

we would like to predict concentrations at a future time period $Z_{t+\ell}$, for $\ell = 1, 2, \ldots$. We had also suggested that the forecast of $Z_{t+\ell}$ would be denoted by $\hat{Z}_t(\ell)$. The function $\hat{Z}_t(\ell)$, $\ell = 1, 2, \ldots$, is known as the "forecast function."

Once a series has been identified, and its parameters estimated for which there are algorithms available, there are techniques available for predicting future values. These techniques are based on certain properties required of the forecast function and have been automated for computer implementation.

The assumption that the random shocks A_t are normally distributed allows us to obtain prediction intervals on our forecasts.

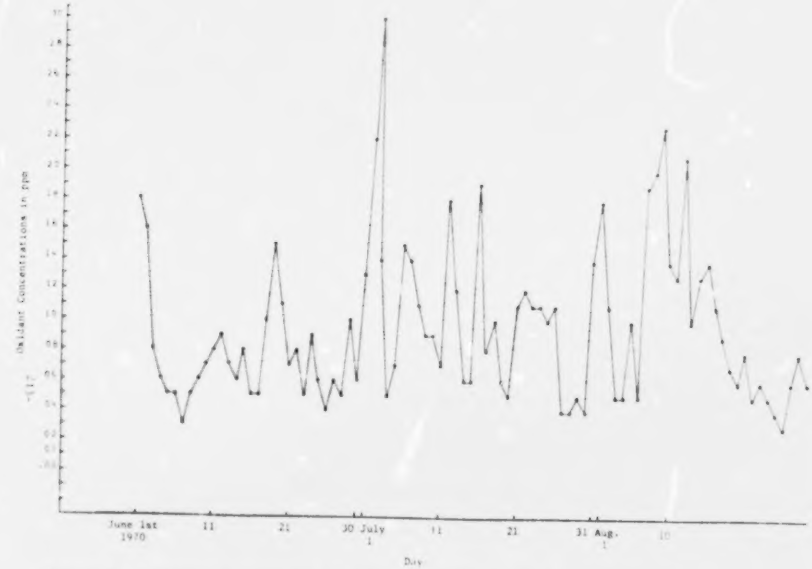
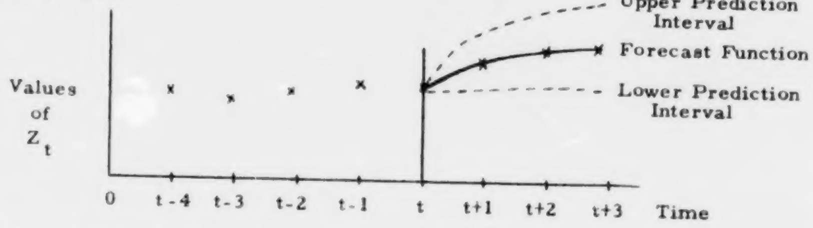


FIGURE 3—*Oxidant Concentrations in ppm for Livermore, California June—August, 1970*

We illustrate these ideas by the following diagram


Upper Prediction Interval
Forecast Function
Lower Prediction Interval
Values of Z_t
0
t-4
t-3
t-2
t-1
t
t+1
t+2
t+3
Time

f) *An Example Illustrating the Use of Time Series Analysis to Air Pollution Data*

Merz et. al. [9] have performed a time series analysis of the air monitoring data from the downtown Los Angeles station. This analysis revealed the existence of certain trends. It was concluded that hydrocarbon, total oxident and carbon monoxide are trending downward. Nitric oxide is trending upward.

Data Base and Method:

The concentration variables were hydrocarbon, oxident carbon monoxide and nitric oxide. The bi-weekly average of the daily maximum hourly average at the downtown Los Angeles station were the data.

A time series analysis of the data covering the period from 1955 to 1967 was made. Using this data a forecast

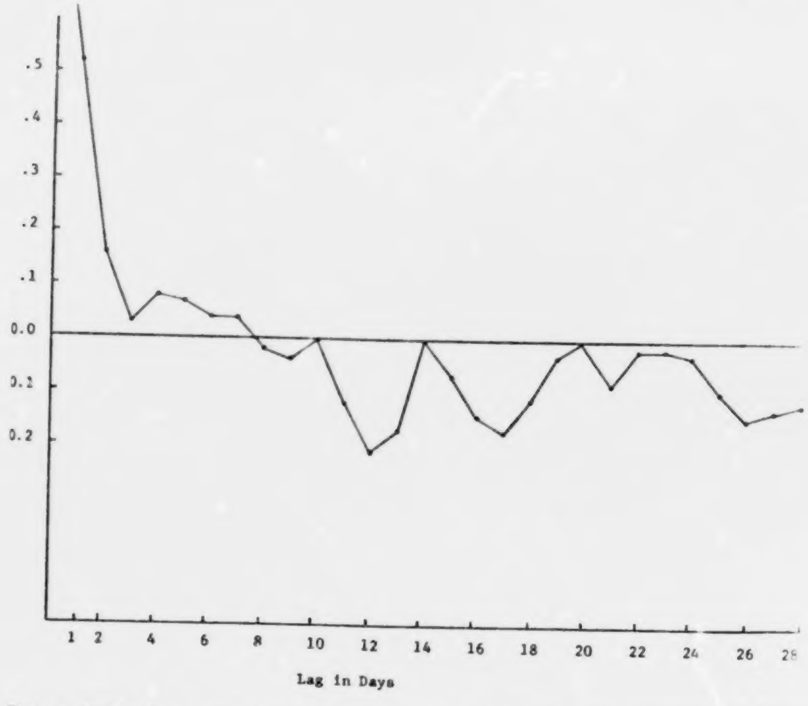


FIGURE 4—*Sample Autocorrelation Function for Oxidant Data From Livermore, California June—August 1970*

was made for the periods covering 1968–1969. A comparison was then made between the forecast and the actual data.

A guiding principle in the choice of models was parsimony; i.e. as few parameters as is possible are used which will still adequately describe the data.

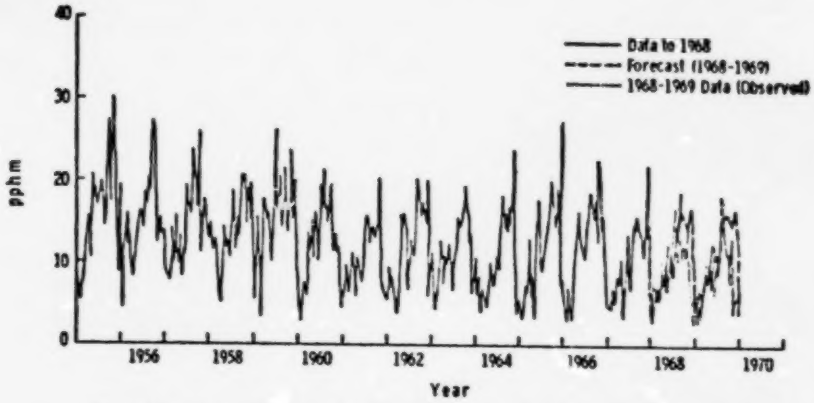
If the forecast rises or falls we have discovered a trend. In this analysis, the random shocks are viewed as related to a combination of variables including meterological factors such as wind speed, humidity, inversion layer, height and light intensity.

Results

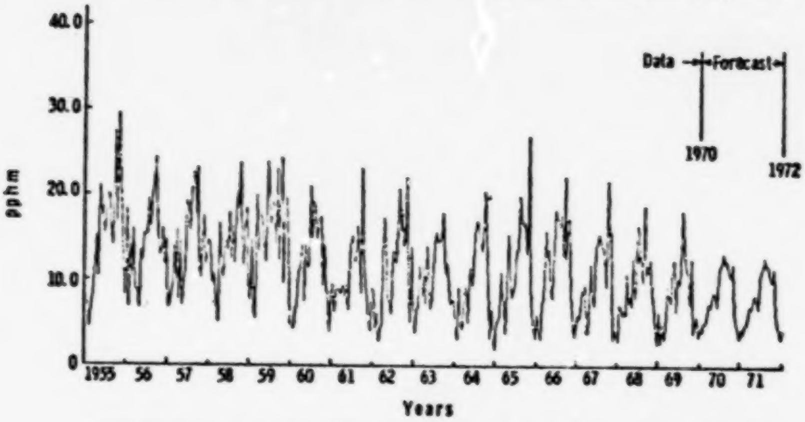
The results in this analysis for the total oxident data are given in the figures below abstracted from Merz et al. [9].

The figures consist of two graphs.

i) a time series analysis and a forecast based on data from 1955 to 1967 together with 1968 and 1969 data
ii) a time series analysis and forecast based on data to the end of 1969



Bi-weekly averages (pphm) of the daily maximum hourly averages for total oxidant, downtown Los Angeles station, data to 1968, forecast to 1970, data for 1968, 1969.



Bi-weekly averages (pphm) of the daily maximum hourly averages for total oxidant, downtown Los Angeles station, data to 1970, forecast to 1972.

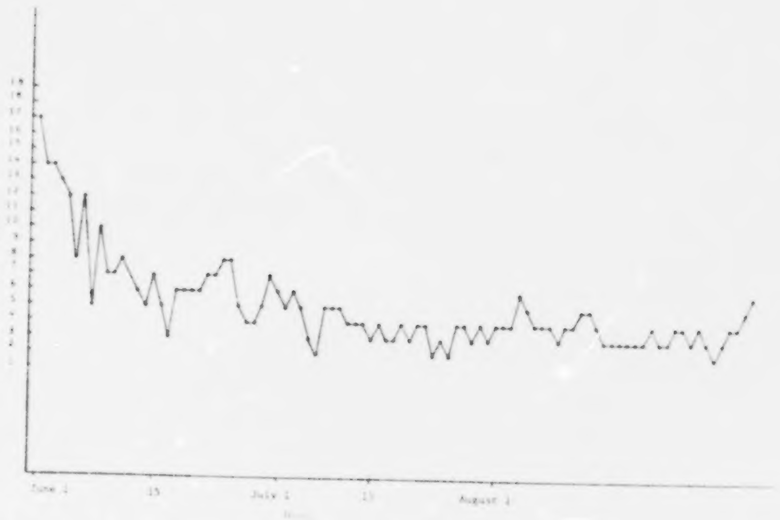


FIGURE 5—*Carbon Monoxide Concentrations in ppm for Livermore, California June—August 1970*

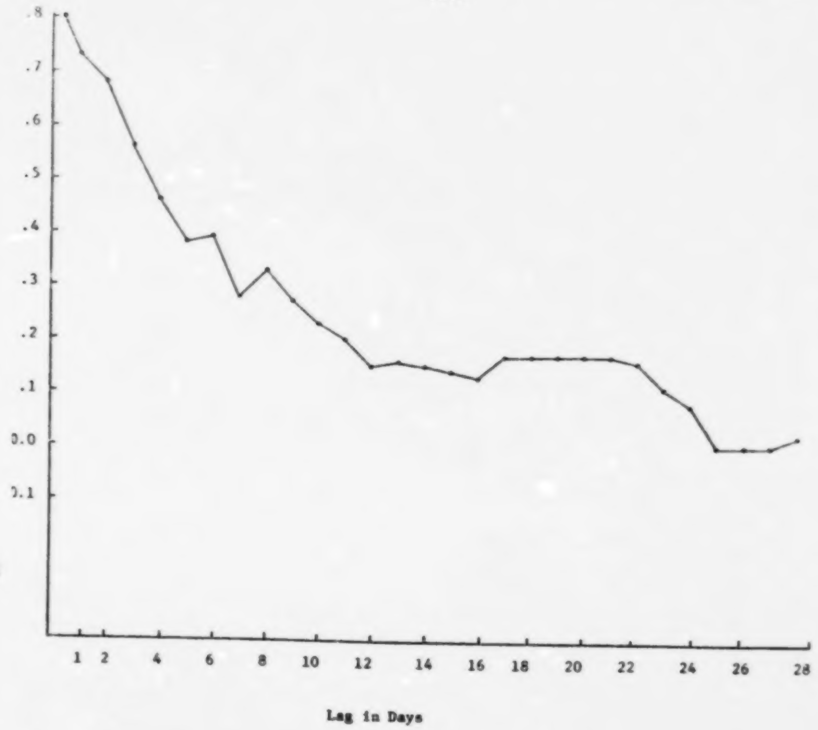


FIGURE 6—*Sample Autocorrelation Function for Carbon Monoxide Data from Livermore, California June—August, 1970*

Comments:

The first analysis and forecast suggested that there would be no change in total oxidant from 1967 on. When the 1968–1969 data were incorporated, a downward trend was suggested. This trend can be ascribed to control measures on the hydrocarbons. 1955–1967 data base: Model ARMA (0,1), with d = 1. 1955–1969 data base: Model ARMA (0,1), with d = 1. The model identified here is a non-stationary model, since d = 1.

B. Deterministic Models

As mentioned before, deterministic models do not have a probabilistic

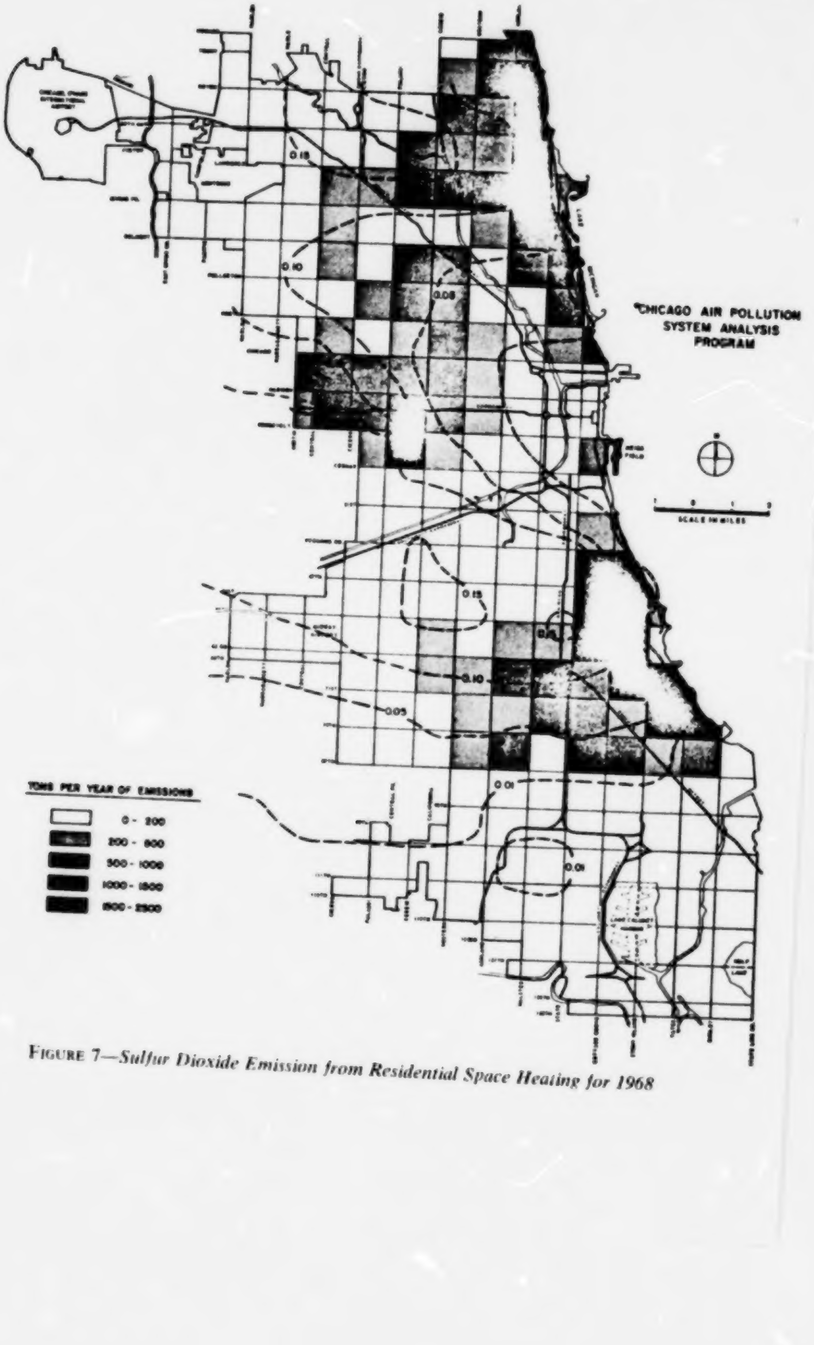

CHICAGO AIR POLLUTION
SYSTEM ANALYSIS
PROGRAM
0.15
0.10
0.05
0.15
0.10
0.05
0.01
0.01
SCALE IN MILES
TONS PER YEAR OF EMISSIONS
0 - 200
200 - 500
500 - 1000
1000 - 1500
1500 - 2500

FIGURE 7—*Sulfur Dioxide Emission from Residential Space Heating for 1968*

structure imposed on it. Prominent among these are the diffusion models proposed by meterologists working on air pollution problems. The other class of deterministic models used in air pollution analysis are the optimization models which will be briefly mentioned.

We shall first discuss a few diffusion models.

1. *Diffusion Models*

Diffusion models serve to gain insight into the relation between meterological elements and air pollution. These models may be likened to a transfer function (a black box) where the input consists of both the combination of weather conditions and the total emission from sources of pollution, and the output is the level of pollutant concentration observed in time and space. These models take into consideration the nature of the sources, the levels of concentration at the receptors, and the atmospheric processes such as photochemical action and eddy diffusions.

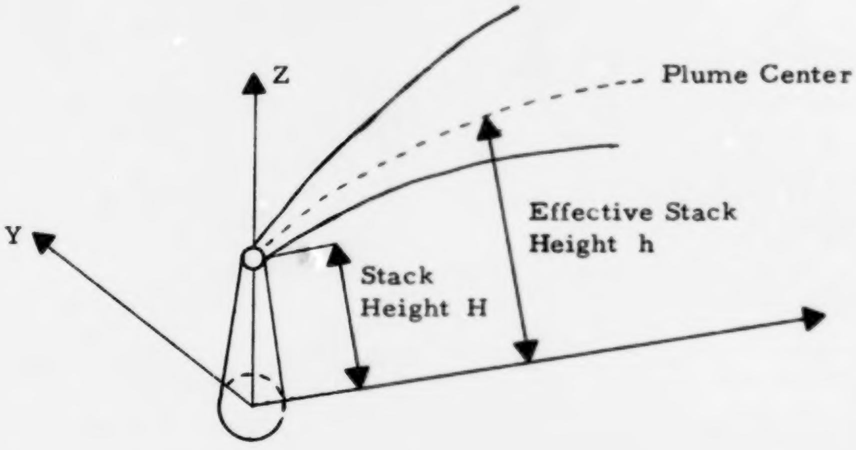
a) *Historical Background*

The pioneer in this work is Frenkiel [13] who developed a mathematical model for the city of Los Angeles. Turner [14] laid the foundation for daily operational models in his development of the Nashville model. One of the most extensive efforts in the development of an urban air pollution model for SO_2 is that of the Argonne National Laboratory [15]. Pooler [16] and Meade and Pasquill [17] have devised models to provide monthly or seasonal values of pollution concentrations.

b) *The Gaussian Plume Equation Model*

Mathematical equations of urban air pollution models describe the process by which pollutants injected into the atmosphere are diluted. The modified Sutton equation, also known as the "Gaussian Plume Equation" enjoys the widest use. Figures 7 and 8 abstracted from "An Urban Atmospheric Dispersion Model" [18] depict the dispersion of pollutants emitted into the atmosphere. Figure 7 shows the distribution over an area, where as Figure 8 shows the propagation of effluents.

Consider an elevated point source as illustrated by a stack shown below. The coordinate system for this stack has its origin at ground level, at the base of the stack. The X-axis extends horizontally in the mean wind direction. The Y-axis is horizontal in the crosswind direction; the Z-axis is vertical.



The "effective stack height" is the height at which the plume center line becomes horizontal—let this be denoted by h.

Let

$X(x,y,z;h)$ = pollutant concentration at point x,y,z for an effective stack height h, $\mu g/m^3$.

Q = emission rate gms/sec.

u = mean wind speed, meters/sec.

σ_y, σ_z = standard deviation of the plume concentration distribution along Y and Z axis respectively. (Note: σ_y and σ_z are functions of downwind distance and atmospheric stability.)

Assumptions:

a) Total reflection of plume at earth's surface.

b) The description of the plume given above represents a description averaged over time; i.e. it is not instantaneous.

c) The plume exhibits a Gaussian distribution of concentration along the y and the z axis.

d) The plume results from a constant emission source.

e) There is no diffusion in the X direction.

Then

$$X(x,y,z;h) = \frac{10^6 Q}{2\pi\sigma_y\sigma_z u}\exp\left[-\frac{1}{2}\left(\frac{y}{\sigma_y}\right)^2\right]\left\{\exp\left[-\frac{1}{2}\left(\frac{z-h}{\sigma_z}\right)^2\right] + \exp\left[-\frac{1}{2}\left(\frac{z+h}{\sigma_z}\right)^2\right]\right\}.$$

Estimates of the ground level concentration are obtained by setting $z = 0$ in the above equation.

Since the wind direction changes continuously over a long term period, for computational purposes discrete wind directions are specified with respect to a 16 point compass corresponding to 22.5 degree circular sectors.

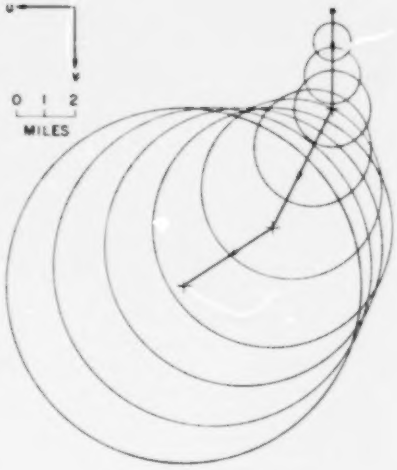


FIGURE 8—*Propagation for Three Hours of an Instantaneous Release Under Assumption* $\sigma_z = \sigma_y$

It is assumed that all wind directions within a given sector occur with equal frequency.

From an examination of the variables, it is clear that several kinds of variables are necessary: wind speed, wind direction, emission rate, σ_y and σ_z.

The values of σ_y and σ_z depend on atmospheric stability which in turn depends upon the variation of temperature with height.

Pasquill has developed Table 4 below for use with Figures 9 and 10 as a means of determining σ_y and σ_z. This table has been abstracted from Turner [14] and Figures 9 and 10.

Calculation of Effective Height of Stack

$$h = H + \Delta H,$$

where

$$\Delta H = \frac{V_s d}{u}\left(1.5 + 2.68 \times 10^{-3} p \frac{T_s - T_a}{T_s} d\right).$$

Table 4. Key to Stability Categories

Surface Wind Speed (at 10 m) m sec⁻¹	Day			Night	
	Incoming Solar Radiation				
	Strong	Moderate	Slight	Thinly Overcast or 4-8 Low Cloud	< 3/8 Cloud
< 2	A	A-B	B		
2-3	A-B	B	C	E	F
3-5	B	B-C	C	D	E
5-6	C	C-D	D	D	D
> 6	C	D	D	D	D

The neutral class, D, should be assumed for overcast conditions during day or night.

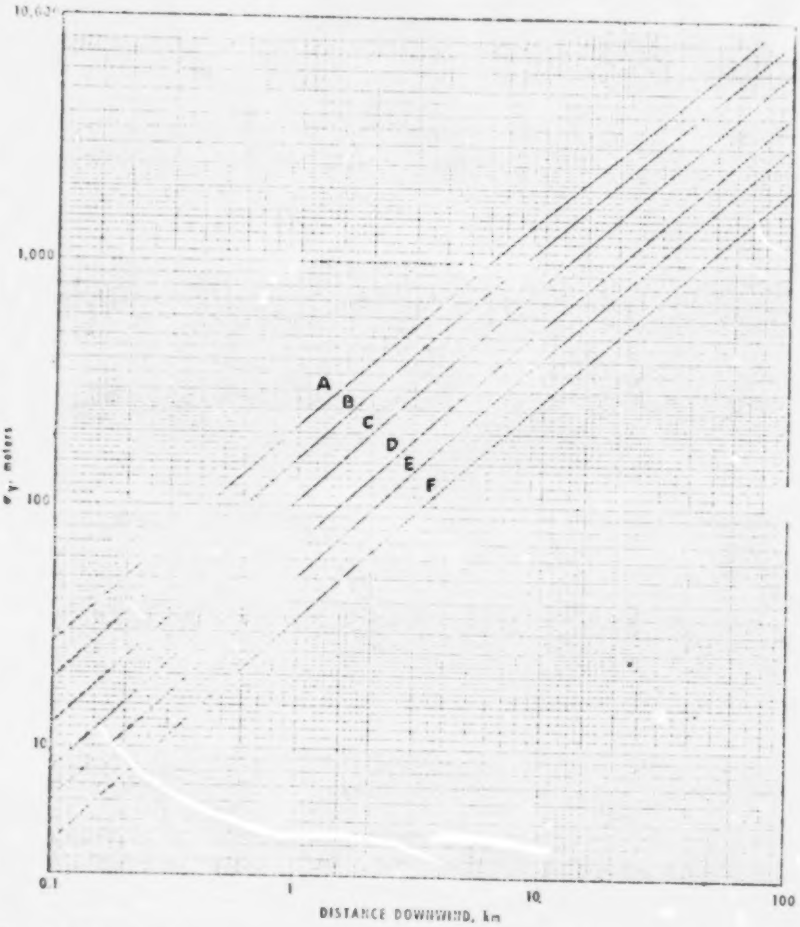


FIGURE 9—*Horizontal dispersion coefficient as a function of downwind distance from the source.*

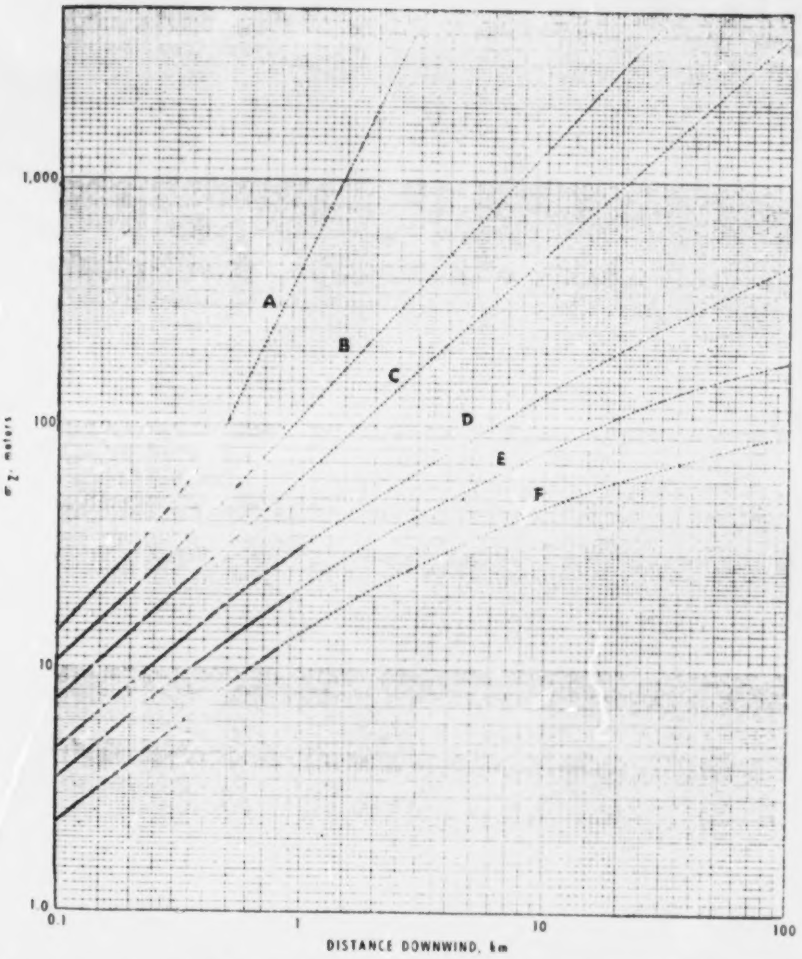


FIGURE 10—*Vertical dispersion coefficient as a function of downwind distance from the source.*

and

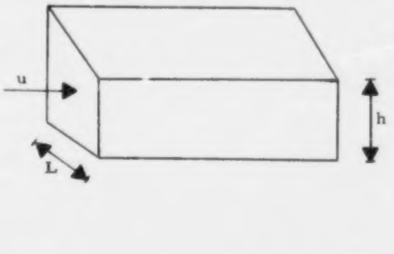
V_s = stack gas exit velocity m/sec.
d = inside stack diameter m.
p = atmospheric pressure mb.
T_s = stack gas temperature °K.
T_a = air temperature °K.

c) *The Box Model*

This model is simpler than the model previously discussed and is used only as a preliminary tool for analysis.

Let

u = mean wind speed
V = volume of box
h = inversion height
ℓ = length of box



Assumptions

a) The source emission rate S (mass/unit time) of any pollutant of interest, into the box is constant with time.

b) There is instantaneous perfect mixing within the box. Thus, the concentration C of any pollutant is uniform throughout the box at any time.

c) The air pressure and temperature are uniform and constant.

d) Reactions involving the pollutants are ignored.

e) The wind blows only clean air into the box at a rate $Q = hLu$. Dirty air of pollutant concentration C, leaves the box at the same rate Q.

Let $C(t)$ = concentration of a pollutant at time t. Then, a mass balance for the pollutant in the box yields the equation

$$V\frac{dc}{dt} = S - QC$$

or

$$\frac{dc}{dt} = \frac{S}{V} - \frac{C}{T}$$

where $T = V/Q$ is the average residence time of a unit mass of the pollutant in the box.

Let C_0 = pollutant concentration at $t = 0$.

Then, $C(t)$ has the solution

$$C(t) = \frac{S}{Q}(1 - e^{-t/T}) + C_0e^{-t/T}$$

and $\lim\limits_{t \to \infty} C(t) = \frac{S}{Q}$, a steady state concentration.

Comments

The assumptions of the model are highly simplified.

- instantaneous mixing does not occur and hence pollutant concentrations are not uniform.
- choice of the average wind velocity is critical.

Example

Let $h = 300$ ft.

$V = 9.3 \times 10^{12}$ cu. ft.

$u = 5$ m.p.h.

$\therefore L = (V/h)^{1/2} = 1.76 \times 10^5$ ft.

$\therefore Q = hLu = 1.39 \times 10^{12}$ ft. 3/hr.

$$T = \frac{V}{Q} = 6.7 \text{ hrs.}$$

It is reasonable to assume that in about one day ($t \approx 4T$) the pollutant concentration will reach its steady state value.

The S for SO_2 is quoted as 396 tons/day or 3.84×10^4 lbs./hr.

$$\therefore C(\infty) = \frac{S}{Q} = 2.76 \times 10^{-8} \text{ lbs./ft.}^3$$

or 4.45×10^{-4} grams/meter.

The air quality standard for SO_2 is 0.5 ppm. or 1.43×10^{-3} grams/meter.

2. *Optimization Models*

The techniques of mathematical programming of which linear and nonlinear programming are the most popular fall under the category of optimization models. Recently, these models have been successfully utilized in several aspects of the decision making and control aspects of air pollution management.

These models along with the implied concepts of viewing complex systems in terms of interrelated activities has proved to be a major contribution to the field of scientific decision making. The versatility and adaptability of these deceptively simple mathematical models appear to have no bounds. In view of the above arguments it is natural that these techniques have found use in air pollution control and management problems.

Despite the proven success of such models in several areas of decision making, their use in air pollution control and management problems appears to be limited. One possible explanation for this may be the fact that up until now analysts involved with mathematical models for air pollution have concentrated more along the lines of data analysis and data management, or have worked along the lines of diffusion and meterological modeling. It is anticipated

Assumptions

a) The source emission rate S (mass/unit time) of any pollutant of interest, into the box is constant with time.
b) There is instantaneous perfect mixing within the box. Thus, the concentration C of any pollutant is uniform throughout the box at any time.
c) The air pressure and temperature are uniform and constant.
d) Reactions involving the pollutants are ignored.
e) The wind blows only clean air into the box at a rate $Q = hLu$. Dirty air of pollutant concentration C, leaves the box at the same rate Q.

Let $C(t)$ = concentration of a pollutant at time t. Then, a mass balance for the pollutant in the box yields the equation

$$V\frac{dc}{dt} = S - QC$$

or

$$\frac{dc}{dt} = \frac{S}{V} - \frac{C}{T}$$

where $T = V/Q$ is the average residence time of a unit mass of the pollutant in the box.

Let C_0 = pollutant concentration at $t = 0$.

Then, $C(t)$ has the solution

$$C(t) = \frac{S}{Q}(1 - e^{-t/T}) + C_0 e^{-t/T}$$

and $\lim\limits_{t \to \infty} C(t) = \frac{S}{Q}$, a steady state concentration.

Comments

The assumptions of the model are highly simplified.

- instantaneous mixing does not occur and hence pollutant concentrations are not uniform.
- choice of the average wind velocity is critical.

Example

Let $h = 300$ ft.
$V = 9.3 \times 10^{12}$ cu. ft.
$u = 5$ m.p.h.
$\therefore L = (V/h)^{1/2} = 1.76 \times 10^{5}$ ft.
$\therefore Q = hLu = 1.39 \times 10^{12}$ ft. 3/hr.
$T = \frac{V}{Q} = 6.7$ hrs.

It is reasonable to assume that in about one day ($t \approx 4T$) the pollutant concentration will reach its steady state value.

The S for SO_2 is quoted as 396 tons/day or 3.84×10^{4} lbs./hr.

$$\therefore C(\infty) = \frac{S}{Q} = 2.76 \times 10^{-8} \text{ lbs./ft.}^3$$

or 4.45×10^{-4} grams/meter.

The air quality standard for SO_2 is 0.5 ppm. or 1.43×10^{-3} grams/meter.

2. *Optimization Models*

The techniques of mathematical programming of which linear and nonlinear programming are the most popular fall under the category of optimization models. Recently, these models have been successfully utilized in several aspects of the decision making and control aspects of air pollution management.

These models along with the implied concepts of viewing complex systems in terms of interrelated activities has proved to be a major contribution to the field of scientific decision making. The versatility and adaptability of these deceptively simple mathematical models appear to have no bounds. In view of the above arguments it is natural that these techniques have found use in air pollution control and management problems.

Despite the proven success of such models in several areas of decision making, their use in air pollution control and management problems appears to be limited. One possible explanation for this may be the fact that up until now analysts involved with mathematical models for air pollution have concentrated more along the lines of data analysis and data management, or have worked along the lines of diffusion and meterological modeling. It is anticipated that once the problems of data analysis and air pollution modeling have been reasonably well solved, decision makers will have to draw upon the optimization models to solve a new generation of problems that will surface. The discussion which follows is written with a view of exposing the decision maker to the capabilities of optimization models in the hope that he will stimulate and encourage their widespread use, whenever feasible.

We shall illustrate below a few such applications.

a) *Application to Alternative Air Pollution Abatement Policies*

In selecting and enforcing air quality standards it is the control authorities responsibility to determine the level of abatement which minimizes for society the total cost of pollution. The total cost of pollution is the cost of damage due to pollution and the cost of pollution control.

A fuel substitution model was developed by A. Teller [19] for estimating the cost of satisfying air quality standards. In this model it is assumed that there are two types of fuel. The objective is to find the minimum cost combination of fuel A and B which satisfies the air quality standard. There are two constraints in the model; these are

i) the combination of high and low polluting fuels should have the same caloric value as before the substitution.
ii) the amount of pollution resulting from this combination should be less than some air quality standard.

The following notations are used.

A = originally used high polluting fuel.
B = substitution fuel—less polluting.
T_{ja} = no. of tons of A presently used by source J.
T_{ja}^{u} = decision variable; no. of tons of fuel A.
T_{jb}^{u} = decision variable; no. of tons of fuel B.
$C_{ja(b)}$ = cost/ton of fuel A(B).
$K_{ja(b)}$ = BTU/ton of fuel A(B).
H_j = no. of BTU's required by source J.
$E_{ja(b)}$ = emission/ton of fuel A(B) from source J.
S_i = air quality standard at receptor I.
M_{ij} = meterological parameter relating emissions at source J to air quality at receptor I.
X_j = percent substitution of fuel B for fuel A at source J.

The linear programming model is

$$\text{Minimize} \sum_{j=1}^{n} (C_{jb} T_{jb}^{u} + C_{ja} T_{ja}^{u})$$

subject to

$$T_{ja}^{u} K_{ja} + T_{jb}^{u} K_{jb} \geq H_j \qquad (j = 1,2, \ldots, n),$$

$$\sum_{j=1}^{n} M_{ij} [T_{ja}^{u} E_{ja} + T_{jb}^{u} E_{jb}] \leq S_i \qquad (i = 1,2, \ldots, m).$$

Computer packages for solving the above model are commercially available.

The objective is to minimize the total cost of fuel for all sources subject to the constraint that for each receptor the sum of the reductions in pollution resulting from each source's use of different fuel is at least as great as the necessary reduction in air pollution concentration.

Assumptions

The above model has the following assumptions inherent to it.

i) fuel substitution is the only method of abatement.
ii) that there can be up to 100% substitution of fuel B for fuel A.
iii) all emissions result from combustion.

Any variations in the above assumptions can be incorporated into the models, though the structure of the above model will slightly change.

The linear programming solution provides much useful information besides obtaining values of the decision variables. This information is outlined in detail by Teller [19].

b) *The Energy Quality Model*

The Energy Quality Model (EQM), discussed in [20], is designed to determine the quantities of fuel, by fuel type, to be supplied to each fuel district to a specified set of demand regions under given assumptions of air quality standards, the demand region's energy requirements and their ability to utilize the fuels and the fuel supply districts' abilities to supply its fuel types.

The EQM is an optimization problem which is designed to distribute fuels from many sources (districts) such that energy requirements of destinations (regions) are met in a manner which minimizes total national cost of obtaining the fuels subject to specified air quality restrictions on emissions and each destination's ability to utilize the fuel types.

The problem discussed here is a standard linear programming problem except for the fact that it has a large number of variables. Such large scale problems can be solved by using what is known as the Dantzig-Wolfe decomposition algorithm, provided that certain conditions are satisfied. Reference [20] provides a detailed method of solving the above stated problem.

References

[1] Larsen, R. I., "A New Mathematical Model of Air Pollutant Concentration Averaging Time and Frequency," *Journal of Air Pollution Control Association*, Vol. 9, pp. 24–30, 1969.

[2] Larsen, R. I., "Air Pollution From Motor Vehicles," *Annals of the New York Academy of Sciences*, 136(12), pp. 275–301, 1966.

[3] Drake, A. W., Kenney, R. L., Morse, P. H., (Editors) *Analysis of Public Systems*, MIT Press, Cambridge, Massachusetts, 1972.

[4] Aitchison, J. and Brown, J. A. C., *The Lognormal Distribution*, Cambridge University Press, 1963.

[5] Barlow, R. E. and Singpurwalla, N. D., "Averging Time and Maxima for Dependent Observations," to appear in the *Proceedings of the Symposium on Statistical Aspects of Air Quality Data*, 1972.

[6] Choi, S. C. and Wette, R., "Maximum Likelihood Estimators of the Parameters of the Gamma Distribution and Their Bias," *Technometrics*, Vol. 11, No. 4, pp. 683–690, 1969.

[7] Wilk, M. P., Gnanadesikan, R. and Huyett, M. J., "Probability Plots for the Gamma Distribution," Technometrics, Vol. 4, pp. 1–20, 1962.

[8] Barlow, R. E., "Averaging Time and Maxima for Air Pollution Concentrations," *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. VI, pp. 433–442, 1972.

[9] Merz, P. H., Painter, L. J., and Ryason, P. R., "Aerometric Data Analysis—Time Series Analysis and Forecast and an Atmospheric Smog Diagram," *Atmospheric Environment*, Vol. 6, pp. 319–342, 1972.

[10] Singpurwalla, N. D., "Extreme Values from a Lognormal Law With Applications to Air Pollution Problems," *Technometrics*, Vol. 14, No. 3, pp. 703–711, 1972.

[11] Gnedenko, B. V., Sur la distribution limité du terme maximum d'une series aleatoire, *Ann. Math.*, Vol. 44, No. 23, 1943.

[12] Esary, J. D., and F. Proschan, "A Reliability Bound for Systems of Maintained Interdependent Components," *Journal of the American Statistical Association*, Vol. 65, pp. 329–338, 1970.

[13] Frenkiel, F. N., "Atmospheric Pollution in Growing Communities," *Smithsonian Report for 1956*, pp. 269–299, 1956.

[14] Turner, D. B., "A Diffusion Model for An Urban Area," *Journal of Applied Meteorology*, Vol. 3, No. 1, pp. 85–91, 1964.

[15] Croke, E. J., Carson, J. E., Gatz, D. F., Moses, H. et. al., "Chicago Air Pollution System Model, Third Quarterly Progress Report," Argonne National Laboratory, ANL/ES–CC–003, 1968.

[16] Pooler, F. Jr., "A Diffusion Model for An Urban Area," *Journal of Applied Meteorlogy*, Vol. 3, No. 1, pp. 85–91, 1964.

[17] Meade, P. J. and Pasquill, T., "A Study of the Average Distribution of Pollution Around Staythorpe," *International Journal of Air and Water Pollution*, Pergamon Press, Vol. 4, Nos. 3/4, pp. 199–211, 1961.

[18] Roberts, J. J., Croke, E. J., and Kennedy, A. S., "An Urban Atmospheric Dispersion Model," Argonne National Laboratory, Report ANL/ES–CC–005, 1969.

[19] Teller, A., "The Use of Linear Programming to Estimate the Cost of Some Alternative Air Pollution Abatement Policies," *Proceedings of the IBM Scientific Computing Symposium on Water and Air Resource Management*, pp. 345–353, IBM, White Plains, New York, 1968.

[20] "Technical Analysis of the Application of the Dantzig-Wolfe Decomposition Technique in the EPA Energy Quality Model," MATHEMATICA, Inc., 1972.

Chapter 4

Models in Water Resources

By

David H. Marks

SUMMARY 105

I. WATER QUALITY 107

A. Description of the Area 107
  1. Problem Identification 107
B. Models in the Aid of Water Quality Decision Making 110
  1. Descriptive Models 111
  2. Management Models 114
C. The Delaware Estuary, An Example 119
D. Summary of Water Quality Models 121

II. WATER QUANTITY 122

A. Description of the Problem 122
  1. Problem Identification 122
  2. Water Use 123
  3. Functional Groupings for Problem Solutions 125
B. Models in the Aid of Water Quantity Decision Making 125
  1. Descriptive Models 125
  2. Management Models 127
  3. Case Study—Rio Colorado, Argentina 129
  4. Summary 133

REFERENCES 133

Models in Water Resources

SUMMARY

This chapter deals with modeling in the aid of decision making for the management of a vital resource: water. The level of complexity and multiplicity of the problems involved in allocating this resource for its desired uses and in limiting the overall impacts of those uses upon the environment and society is great. We find it convenient to categorize our discussion by considering the two areas of water quality and water quantity. Although there are interactions between these two areas, the division nevertheless helps to focus on two different aspects of a similar problem: optimal resource allocation in a public sector problems where, unlike the private sector, there is no complete pricing mechanism to guide the decision or express preferences. The focus on water quality is pollution which we define as waste inputs to water which preclude or damage the employment of water for other uses such as recreation, water supply or wild life. Pollution loads are produced both by the disposition of residuals from process use (industrial and municipal wastes), and as a by-product of development and land use policies (decreased quality of urban storm runoff and increased sediment loads). Are there better uses for the residuals, better development policies than the present mechanisms or is better quality justified to support other uses? As with any public sector decision problem, the costs and benefits of any action to bring about improvement are not all directly measurable in commensurate units or towards the same interest groups. Thus the decision maker must find some formal mechanism to help balance these diverse factors in the choice of when, where, how and how much should be invested for water quality.

In water quantity, we are concerned with the physical aspects of water and its allocation to proper use. The availability of water is spread quite differently spatially and temporally than where it is needed. Thus the decision maker is concerned with designing investments and strategies for capturing and transporting the resource in reliable quantity. Occasionally the problem is the control of excessive water as in flooding. Decisions as to what these uses should be and how much should be allocated are again of great importance. There are diverse interest groups with different preferences and objectives; costs and benefits are not in commensurate units and may fall differently on the various segments of society. The decision maker again must weigh and balance these factors in making investment decisions.

The role of models in the study of these decisions is pervasive. There are many different aspects of the problem where analysis using modeling is the principal information input to the decision process. In this chapter modeling in water resources is discussed from the perspective of the problems addressed and where new developments can help improve decision making.

Water Quality

The decision makers charged with deciding how much, where and when investment should be made and what strategy should be used require information about a variety of questions to define and evaluate the options available. Models to aid in this process fall

Preceding page blank

into two categories: descriptive and management. Descriptive models show the impact of changing pollutional patterns. The most prevalent models of this sort are those that show the response in space and time of a water body to inputs of pollutants such as municipal sewage or heated water. Normally the response is measured in terms of some physical parameter of the water body such as its dissolved oxygen concentration or temperature. Less well developed are models that translate these physical parameters into estimates of damages or benefits. Ecosystem modeling has received attention in the past few years but is still in its infancy. Models to show economic, demographic and social impacts on water quality changes and thus estimate benefits and costs have also been developed. Other processes which produce pollutants such as storm runoff have been modeled, as well as the technologic components for controlling pollution.

Management models are developed to help choose from among alternative control strategies those which best satisfy given quality goals. Generally these models focus on the decision to be made by a regional authority for investment in and implementation of a water quality management plan.

Water Quantity

Here the decision maker must decide on how water is obtained, controlled and transported and to what use it should be put. Descriptive models help to evaluate how much of the stochastically varying resource will be available and how storage or other policy decisions will operate to capture the resource and/or to reduce its flooding impacts. Management models help to investigate investment decisions in control, storage and transport, as well as which allocation of funds to use by selecting from the available options that set which best satisfies objectives. Here again these models must formally identify the multi-objective nature of the problem and account for costs and benefits by the appropriate objectives in order to provide important information to the decision process.

The water quality and quantity models discussed in this chapter are given in the Summary Table.

WATER RESOURCES

Models Discussed in Chapter

Summary Table

Model/Decision Area	General Type	Important Characteristics
Water Quality		
Pollution Impacts in Water Bodies	Analytic Simulation Statistics	Shows impacts on physical parameters of water bodies due to waste inputs (usual parameters dissolved oxygen, nutrients, salinity and temperature)
Ecologic Modeling	Simulation	Models address broad strategy of organisms and specific tactics
Urbanized Storm Runoff Modeling	Simulation Statistics	Shows quality aspects of urban runoff and allows both land use and structural controls
Models of Treatment Technology	Simulation	Helps to design, size and cost treatment facility
Models of the Economic and Social Impacts of Water Pollution Control	Input-Output Analysis Simulation Systems Analysis Games Statistics	Attempts at defining economic and demographic charges and to aid in implementation
Pollution Impacts on Ground Water	Simulation	Temporal and spatial dispersion of pollutants

Model Decision Area	General Type	Important Characteristics
Regional Waste Management	Optimization (linear, dynamic, integer and geometric programming)	Investment models for the following problems: Regional least cost treatment, equitable cost distribution, treatment plant location, low flow augmentation, in-stream aeration, bottom deposits, storm water
Optimal Taxing	Optimization (linear and nonlinear programming)	Tax schemes to support regional investment
Power Plant Siting	Optimization (linear programming)	Regional supply models which consider environmental standards
Watershed and Catchment Models	Simulation	Shows the response in terms of runoff to a rainfall intensity and pattern
Synthetic Rainfall and Stream Flow Generation	Statistics	Generates long synthetic rainfall and stream flow data traces which have similar statistical properties as short observations.
Groundwater Modeling	Simulation	Movement and availability of ground water
Economic and Social Impacts of Water Resource Development	Simulation Systems Dynamics Statistics	Models for predicting economic, demographic recreational effects of projects
Regional Water Demand Projections	Input-output Modeling	Estimates water requirements in various use sectors and categories
Regional Water Supply Analysis	Linear Programming	Finds optimal supply investment for a given demand
Preference Analysis and Implementation	Game Theory and Optimization	
Reservoir Management	Statistics Simulation Optimization	Regulating operation to provide a stochastically varying resource to varying competing demands.
Investment Scheduling and Sequencing	Optimization	Project scheduling under budget constraints
Pipe and Aqueduct Networks Design	Optimization Simulation	Sizing and configuration analysis
Gauging Station Location	Optimization Simulation Statistics	Gauging location to maximize information
Technology Design	Optimization (dynamic programming, geometric programming)	Optimal sizing and choice of subsystems for treatment

I. WATER QUALITY

A. Description of the Area

1. Problem Identification

In order to consider the variety of models helpful in the decision processes for establishing the management of water quality, it is necessary to focus on the factors that cause the problem, what type of decisions can be made to deal with the problem and who will make them.

What Causes the Problem—Water quality problems are caused by the use of water as a disposal medium to the extent that other uses of the resources are impaired or foregone. We must view the assimilative capacity of water to receive and degrade wastes as a valuable but limited resource that must be properly allocated just like any other resource. Unfortunately, the traditional private sector pricing mechanism which expresses society's preferences for other types of resource use breaks down in the case of a common, public resource such as the ability of the environment to assimilate wastes. The result has been an overuse of the environmental resource leading to its deterioration with little attempt to recover and reuse waste material. A brief recounting of the type of processes that create wastes which eventually enter water and the other water uses which waste disposal may interfere with serves to highlight the diverse, complex nature of the problem.

It is possible to differentiate between two different ways that waste materials enter water. This distinction is important for it helps to recognize those waste sources that are more amenable to direct control. The first way is through water use. There are many different water uses and most of these uses return water to water bodies that is diminished in quantity and quality. Sometimes the addition of wastes to water is deliberate in that the water is being used as a transport mechanism to take wastes away from the place where they are generated. In other cases, the water is being used for a particular process and becomes contaminated by materials it come in contact with. In both of these cases, the source of the waste and the point at which it enters the water body is specific and identifiable. One can think of a variety of control alternatives ranging from technological devices located "at the end of the pipe", to process changes and other policies which will either lessen water use or change the wastes that enter the stream. These are referred to as point source wastes. Examples of point source wastes are those from domestic usage, industrial use and heated water discharges from cooling usage.

Another type of waste enters water bodies as the result of a combination of the normal rainfall runoff process and policies for the disposition of land and air resources. These include the pollution of surface runoff to streams and to groundwater from non-urban areas and agricultural areas. Such sources are non-point source problems and are normally amenable only to changes in policies and to process changes which keep wastes away from the runoff process. There is little large scale technology that is helpful. Sources include storm water runoff to water bodies from urban and agricultural areas.

The primary uses of water which have a major quality concern are municipal, industrial, cooling, recreation, aesthetics, navigation, agriculture, and waste disposal. It is clear that there is conflict between these uses in that each require a different level of quality and each deteriorates quality through use in a different manner. Certainly a legitimate use for water is the disposal of wastes, but it can be shown that this use in excess precludes other uses. Thus waste discharges to a water body only become pollutants when they interfere with some other desired use. If it is decided to use a stream mainly for waste disposal and no other use is desired except navigation, this would not in the technical sense be pollution. However, such single object use of water is generally deemed socially unacceptable in the U.S. and an attempt is made to handle more than one use for a water body. Still, requirements to use the water for non-contact recreation, industrial processes and cooling may be far lower than using the water for water contact recreation, fish and wildlife and municipal water supply. Thus different levels of improved water use may be chosen.

What Decisions Can Be Made?

As with any scarce common resource problem, the public sector must intervene to ensure that resource allocation does reflect the objectives of society. This intervention can come in the form of laws that prohibit certain types of activities that create pollution, require waste treatment and set environmental standards, or in economic incentives such as taxes or subsidies which act as price surrogates to cause the private sector to shift its resource use. It is clear that to establish rational policies for the limiting of such environmental deterioration, a careful weighting of all the complex technical, economic and social implications of waste management upon the diverse and often conflicting objectives of society is of the utmost importance. For instance, every process that develops goods and services for our basic well-being, as well as the extra trappings of an affluent society, requires the use of limited resources and produces residuals which must be accounted for in some manner. We must consider how much of these residuals will be tolerated in the environment when the alternatives to reduce these may mean serious economic costs in resources, lost jobs or even the elimination of certain current development activities. A simple law requiring pollution to stop does not take into account the objectives and purposes of the processes that created the pollution in the first place. Thus, the decision maker, who must balance the answers to these delicate questions, is faced with a very complex, difficult and quite subjective task. This chapter is devoted to the role that analytical models have played and can play in ordering policy decisions, both by helping to quantify some of the trade-offs between objectives for the system and by providing better insight into how the system operates. The questions the decision maker must face center around two broad but interactive issues. What quality is desired and what is the best way to obtain it? These questions help to bring to the forefront the prominent decision issues. It is clear that there is no answer to one question without actually investigating both simultaneously. The quality desired is very much a function of how much that quality will cost, not only in direct economic costs, but in terms of opportunities foregone for investment in other worthy public sector activities such as hospitals, roads, schools and parks. As with any public sector problem, the recipients of the benefits of the improvements are diffuse and, due to the complexities of the system, we are not capable of assessing directly the value received. Thus, the question of how costs will be allocated and whether that allocation is equitable often turns out to be a more pertinent question than what the total cost is. Further, since the goal of the decision process is to come up with a plan that will and can be implementable, care must be taken to see that the participants in the plan understand and accept its objectives and equity, and that the proper institutions exist for running the plan and assessing the costs.

The problem of establishing quality standards is extremely difficult for it requires the ability to predict the cause and effect relationship between waste discharges and water quality, not only in terms of physical impacts on other water uses, but on their social and economic reverberations as well. This has led to considerable use of models to identify such cause and effect relationships and for evaluating control alternatives.

The main difficulty for the technical investigations will be in defining the quality that is desired. Without the ability to assess preferences for water quality and water uses for all users, we must fall back on attempts to define possible water uses and their cost implications.

The difficulty in determining what water use to choose falls into two separate categories: determining a set of physical parameters that show if a certain water use is possible or not, and determining whether a particular water use is justifiable within a region in terms of impacts on economic and social objectives. It is clear that the second task is very difficult, including as it does assumptions about social preference and the linkage between water quality and regional activity. However, it turns out that the first problem, that of finding physical parameters, is a major definitional problem with which the various standard-setting agencies have to deal. Because most organic material degrades in natural water bodies by microorganisms which use dissolved oxygen in the stream for the degradation process, stream dissolved oxygen (D.O.) is the most commonly used surrogate indicator for overall water quality. However, this by itself is not a good indicator for all situations. For municipal water supply, there is a concern for chemicals and pathogenic organisms concentrations. For cooling water the most important characteristic is the temperature. Thus, choosing the desired water quality indicator is an exercise in identifying the important pollutants in the area and how they reflect on water quality in the water body. Also, concern for the types of water use possible in the area and the improvements necessary to bring about that quality must be considered. The key problem is to then balance the cost of eliminating to some degree the waste sources against the benefits of the resulting quality increase. As quality increases are often implicitly stated, this balancing is a subjective process, as will be shown in the discussion of the Delaware Estuary model.

In general, the role of models in the process is one of clarification, not decision making. A variety of models increases the decision maker's understanding of his options and the sensitivity of his policy decisions.

Who Makes the Decision

The public sector nature of the problem requires the intervention of government to order the many diverse interest groups. The term decision maker, while quite comforting to the analyst who likes to think that one person assimilates all the information and decides, in fact is a euphemism for a very diffuse decentralized coalition of pressure groups and different levels of governmental officials.

In the U.S., the major role of establishing water quality standards and types of waste control that will meet those standards fall to Federal, State and regional organizations. The regional management organization is a recent attempt at problem oriented institutions that have legal authority to manage all aspects of the problem without questions of state and local political boundaries. An example is the Delaware River Basin Commission which has control over all water resources activities related to the Delaware River in the states of New York, New Jersey, Pennsylvania and Delaware. It represents a partnership of these states and the federal government to implement control strategies and quality plans. Such innovative regional organization will continue to provide major impetus for quality control. Most management modeling is done from the perspective of the regional authority as the decision maker.

The major lead on setting broad scale policy questions and in generating money for improvements has been the federal government through the Environmental Protection Agency. As a participant, as well, in the regional organizations it exerts a powerful influence. The role of the States has been reduced to that of participating in regional schemes and otherwise implementing national goals.

B. Models in the Aid of Water Quality Decision Making

The models used in water quality decisions are closely related to the types of decisions being made which are in turn influenced by the character and nature of the problem. In most problems there are two types of models that can be considered. One is a descriptive model which explains the cause and effect relationship between a particular policy and the important parameters in the system. Most simulation models fall into this category. The important aspect here is that the model is descriptive: it shows the results of a given policy. The second type of model is the management model which is both descriptive and prescriptive in that it shows cause and effect and attempts to choose between alternatives which best meet established objectives and constraints. Most optimization models are of this form. Both types of models have their individual defects. While optimization models can choose between alternatives, structural considerations limit one to very simplastic representations of the underlying cause and effect relationships. On the other hand, while descriptive models can be made as detailed and as intricate as necessary, they can consider only one alternative at a time. Thus in most applications such methods are used in consort; the management model suggests areas of good design for further consideration by the descriptive models.

1. *Descriptive Models*

Waste Input—Physical System Response Models. From the earliest investigations in water quality the engineers have built mathematical and analog models to represent the response of the physical system to waste inputs. This allowed the investigation of the response of the physical system when new control plans or new waste inputs were put into action. Prominent among these models are *dissolved oxygen models* which show the biological response of the water body to organic inputs; *temperature models* which show the temperature response to heated water inputs; and *salinity models* which measure the amount of salinity intrusion in estuaries as a function of the fresh water flow in the estuary. We next discuss each in turn.

The *modeling of dissolved oxygen* has evolved from the earliest case of an assumed one-dimensional, steady-state stream effected only by carbonaceous wastes to the present state-of-the-art which allows consideration of both spacial and temporal considerations in stream and tidal estuaries for both carbonaceous and nitrogenous wastes. An excellent reference is the Tracor Report (104) which was prepared for EPA as an assessment of the art in estuarine modeling. Estuaries, which are the sea-tidal portion of large rivers, are of particular interest as they are usually most heavily populated and industrialized. Further, they are biologically important because they are major habitat production areas. Finally, from the hydrodynamic point of view, they represent complex analysis problems because of the interaction of the ocean tides and the incoming fresh water inflow from upstream. The Tracor Report deals with three- and one-dimensional temporal models of estuary hydrodynamics, and includes the consideration of dissolved oxygen effects caused by carbonaceous and nitrogenous wastes for the one-dimension, steady-state, estuary case; the case most susceptible to large-scale solution. Solution of the differential equation model used to represent the process is accomplished by analytical integration of the differential equations over time. For anything but the most drastic simplifying assumptions, this proves to be very difficult. From a large-scale, computational point of view, a finite section approach is more feasible. The estuary is divided into a series of segments and it is assumed that each section is completely mixed. This in essence replaces the derivatives in the steady-state equation with difference approximations. A one-dimensional, time-varying model for a two-stage, Biochemical Oxygen Demand—Dissolved Oxygen (BOD-DO) reaction which translates organic inputs into a distribution of organics in the estuary (Equation A1), then translates those organics into estuary dissolved oxygen (Equation A2) would take the following form:

$$\frac{\partial L}{\partial t} = -\frac{1}{A}\frac{\partial}{\partial x}(QL) + \frac{1}{A}\frac{\partial}{\partial x}\left(EA\frac{\partial L}{\partial x}\right) - K_d L \qquad (A\,1)$$

$$\frac{\partial c}{\partial t} = -\frac{1}{A}\frac{\partial}{\partial x}(Qc) + \frac{1}{A}\frac{\partial}{\partial x}\left(EA\frac{\partial c}{\partial x}\right) + K_a(c_s - c) - K_d L + P - R - B \qquad (A\,2)$$

where L = a measure of the organic material in the estuary due to waste loadings

K_a = reaeration coefficient

K_d = decay coefficient for the reduction of stream organics by biologic action

P = photosynthetic oxygen inputs

R = losses in oxygen due to biologic respiration

B = losses in oxygen due to decomposition of bottom deposits
A = the cross-sectional area of the estuary
Q = the freshwater inflow
E = dispersion coefficient (tidal effect)
t = time coordinate
x = distance coordinate measured along the estuary
c = dissolved oxygen concentration in the system
c_s = dissolved oxygen concentration at saturation

The inclusion of temporal variations of this form are accomplished using a difference method with the equations being stepped through in time. Such a method is described in Pence et. al. (79) and is included in a model distributed by the EPA (30).

For non-estuarine rivers, EPA has also been instrumental in establishing both steady state and time varying models of dissolved oxygen behavior. They have aided state agencies in the implementation of two models developed for the Texas Water Plan which are available in computer packages. DOSAG-1 (101) is a steady state model which includes both carbonaceous and nitrogenous leads and may be operated on to find stream D.O. response to several different sets of input stream and waste loading conditions. The model will also calculate the amount of water needed for flow augmentation to reach a specified dissolved oxygen standard. QUAL-1, a time varying non-steady state model (Texas 100, 102) allows the investigation of stream dissolved oxygen on an hourly basis when input parameters are varied in that time scale. This type of model is used to determine how often and to what magnitude stream conditions vary over time due to a given control policy.

We next turn our attention to *stream temperature models*. With electrical energy production already being a major withdrawal user of water and with anticipated future demands, the question of how heated discharges impact on receiving waters is one of great importance. Every new power generation site must be carefully evaluated to show how its discharge will be distributed, not only for ecological protection, but to insure that the plant's efficiency is not impaired by the heating of its cooling water. Extensive modeling has been done for the problem of translating heated discharges to stream, estuary and coastal site water temperature profiles. Examples can be found in the Tracor Report (104) and in Stolzenbach, Adams and Harlemen (95) of the state-of-the-art in modeling such discharges. The successful translation of these temperature increases to estimates of biological change is still very much in its infancy and is addressed in another section.

The last set of models within the general framework of physical system response models deals with *salinity intrusion*. This area is of interest in estuarine waters because salinity in water can destroy its use for most municipal and industrial processes and because it can mean rapid change for biological communities. Salinity intrusion is most strongly retarded by the magnitude of the fresh-water flow in the river portion of the estuary. In summer and drought periods, the lessening of this flow can mean serious increases in salinity at points where fresh water is normally derived for other purposes. Models of this process are helpful in planning for emergencies and for evaluation of the effect of upstream water diversions on downstream salinity. The model developed by Thatcher and Harleman (103) is an example of such a model.

Ecologic Modeling. The term ecologic modeling will be used to describe those models which attempt to predict the changes in ecologic communities due to changes in waste control policies. As Holling (47) points out, the use of mathematical modeling and in particular computer simulation in this area has taken two main thrusts: one that concerns itself with the broad strategies used by organisms and the other that is concerned about the specific tactics used by organisms. The first leads to consideration of ecosystem organization and stability in terms of this evolution of populations and of communities, while the second deals with analysis of the mechanisms and underlying interaction between parts of ecological systems. These models are described in detail in Pattern (79) and O'Neill et al. (77). It should be noted however, that the art is still not to the point where direct ecologic evaluation of the benefits (positive or negative) of a change in water control policy for a specific area can be made. However, the amount of attention and excellent work being devoted in this area is very encouraging for future developments.

Urbanized Storm Runoff Modeling. As noted earlier, runoff from storms in urbanized areas picks up waste material as it is collected in storm sewer systems and then discharged to receiving waters. The problem of how to control such pollutional loads is strongly amplified by the fact that the waste flows occur only a limited number of times during a year and then have very large volumes in short time spans. Developing treatment for such infrequent high volume waste flows is virtually impossible without some consideration of storage and other means of slowing down and spreading out the loads. The most important model developed in this area is called the EPA Stormwater Model and it is described in Lager, Pyatt and Shubinski (60). For a detailed representation of an urban catchment including population estimates, configuration of the sewer and storage system, time since last storm and a given design storm, the model develops, not only the hydrographs of the volume of water moving at any time and point through the system, but its quality as well. Initial investigation shows the water quality of storm runoff to be worst in the earliest part of the storm thus allowing a strategy of treating and storing only that portion. This important model in effect allows the evaluation of proposed combined storage and treatment facilities, as well as the pollutional impact of storm water collection systems. It is routinely used by consultants and local governments, as well as higher levels in the decision making process involved in the control and treatment of urban storm drainage. Other models for clarifying land use policy and water quality interactions are discussed in Jameson (53).

Models of Treatment Technology. The technology for treatment of municipal waste discharges represents a variety of different physical, chemical and biological components, which must be sized and developed within a total systems context. A major modeling effort by EPA in developing a computer system for sizing and evaluating waste treatment systems has been developed by Smith (93) and has proven very helpful in estimating general treatment costs and trying out design alternatives.

Models of the Economic and Social Impacts of Water Pollution Control. From the point of view of evaluation of control alternatives to decide what level of quality is desired, some evidence of the benefits of such a policy is necessary. We would like to know, for instance, what impacts, plus or minus, a river clean-up would have upon jobs, economic activity, increased recreation participation and other gross indicators of social well-being. The history of modeling in such questions is at best fragmentary and far behind our ability to model the physical system response. This is not surprising as the physical system is much better defined and understood. For the most part, social system benefit estimates have been based on intuitive models or simple regression estimates. Recent work indicates that analytic modeling can be carried out in this sector. Use of regional input-output tables specially configured for calculating the economic impacts of pollution control is shown by Miernyk (75). Use of systems dynamics to show the feedback interaction between water resource investment and demographic and economic trends in a region has been developed by Hamilton et al. (42). Gates et al. (33) also used systems dynamics to show political and institutional response to new water quality legislation in California. Attempts by the EPA to develop simple models to be used in simulated games representing the decision to determine preferences and strategies for reaching implementation are reported by House and Patterson (48). Such attempts are very encouraging and point to a much better understanding of the evaluation of all the impacts of governmental resource development in the face of a serious conflict in objectives.

Models of Pollution Impact on Groundwater. Models to show the impact on physical parameters due to pollutional inputs have reached a fairly high level of sophistication where both temporal and spatial effects can be noted. The article by Bradehoeft and Pinder (9) and Reddel and Sunda (82) are examples of the computational and theoretic state of the art in this area.

2. *Management Models*

Here we begin an investigation of models which allow a selection between alternatives according to specified objectives and constraints. Such models are most often used in consort with the descriptive models such as those described in the previous section. The reason for this is that the management model must of necessity have a more simplistic view of the cause and effect relationships in the system, and the descriptive models are helpful in checking out in more detailed fashion the system configurations suggested by the management models. There are issues introduced when considering management models which have to this point been rather subjectively hidden in the discussion of the descriptive models. This is the choice of objectives and the measures of effectiveness for those objectives necessary for evaluation of a solution. Optimization techniques usually require a one-dimensional objective in order to be effective, while any problem worth its analysis, particularly in water quality management, has many diverse and conflicting objectives. The approach used in almost all of the management models to handle multi-objectives is a surrogate approach which allows the use of single objective optimization techniques. One objective used is national income (economic efficiency). This is normally not measured by net benefits since pollution control benefits for this objective are difficult to identify in economic terms except by the cost of the control procedures. This in most cases is represented in the objective function. Another important objective is society's desires for environment quality. This is represented in the model by specified quality standard constraints which must be met. Another issue is the equity of the solution and its ease of administration. This too is also handled as a constraint but in a more subjective manner and is discussed in detail later in this section.

Our discussion of the management models is best focused by looking at particular types of problems. One major problem area, with many submodels and much research, is the range of choice open in the investigation of water quality in complex regional urban settings. Another problem, usually considered quite independently, is the energy expansion-environmental effects problem faced in the location of new electrical generation facilities. Management models in optimal technology design are also discussed.

Regional Waste Management Problems. In any complex urban setting the combination of large population and industrial enterprise means that there will be serious water quality problems that require control because of impacts on other water uses. The specific problems that contribute to this general deterioration of quality include: municipal waste discharges, industrial waste discharges, urban storm drainage, tributary and upstream loading and bottom deposits from previous pollutional loadings. While we have outlined control schemes for most of the individual pollutional sources, it is clear that the best alternatives, particularly in the case of runoff and other non-point sources of pollution, may well be through land use or other policy changes that keep wastes out of water in the first place. Also, when one begins to look at the region as a whole, there appears to be a reasonable basis for looking for economies of scale in control through regional cooperation in the areas of joint treatment facilities, measures for increasing the assimilative capacity of the water body and better land use planning.

The region as a whole seems a good way of defining an administrative unit within which the costs of control can be successfully allocated among the people in the region. Thus, the problem referred to as the regional management problem is the problem of finding the set of policies, technology and control schemes that satisfies the regional desires for environmental quality and equitable implementation at least cost. Ideally, such a model to sort through all these alternatives would allow as decision variables all possible control and policy options. This in terms of computational and conceptual problems has not been possible. Modeling instead has been directed towards the investigations of a set of sub-problems within the region to supply information about the options available. These models include the regional effluent control model, the regional treatment facility location problem, flow augmentation by upstream releases, instream aeration for increasing the assimilative capacity of the water body, urban storm water management, nontreatment control alternatives such as the removal of bottom deposits and other problems. These models are directed at developing a scheme that would then have to be administered by some form of regional authority and do not explicitly consider the implementation problem. In another field of work, many investigators have looked for pricing schemes that would allow a decentralization of the implementation process. Through the imposition of charges, polluters could be made to adjust their waste loads to a more socially desirable alternative. The problem comes in deciding what structure and magnitude such pricing schemes should take to be implementable.

The problem of regional effluent control was the first regional model attempted and is directed towards finding the set of effluent changes at municipal and industrial sources necessary to reach desired quality standards measured in terms of stream dissolved at specific points along the water body. Many authors have suggested models for this problem which essentially take the form of a mathematical programming problem with linear constraints:

$$\text{Minimize } z = \sum_j f_j(x_j) \qquad \text{(A3)}$$

$$\text{Subject to } \quad \sum_j A_{ij}(x_j) \geq b_i$$

$$i = 1, \ldots, m \qquad \text{(A4)}$$

$$0 \leq x_j \leq u_j$$

$$j = 1, \ldots, n \qquad \text{(A5)}$$

where x_j = the amount of waste to be removed at source j in pounds per day of Biochemical Oxygen Demand (decision variable—unknown).

u_j = the upper bound on waste that can be removed at source j in pounds per day of Biochemical Oxygen Demand (known).

$f_j(x_j)$ = the cost of removal of x_j (known).

$A_{ij}(x_j)$ = the quality response at some point i in the water body caused by a waste loading of x_j at source j, for a given set of physical parameters (known).

b_i = the minimum desired improvement in quality at point i (known).

There are several important assumptions buried in this simple formulation that should be discussed. Notice that the objective (A3) calls for the minimization of the cost to the region as a whole without any notion of the allocation of the cost. The choice of which sources will be reduced will be based on a combination of strategic location and efficient methods for treating wastes, rather than on some notion of how much damage is being caused by the waste source. The formulation stresses an overall efficiency objective without taking into account distributional problems. Thus it would be possible, and has in fact occurred in several application studies, that large, inefficient sources which have done little to concentrate or reduce their waste flows are required to make no changes even though their damage is great, because of the great cost involved in redoing the whole plant. Meanwhile, small, highly efficient sources where investment has already been made will be called on to treat further, since wastes are already concentrated. This calls for some means to reallocate the efficiency costs of the improvement either among the polluters or among the beneficiaries, and some regional institution to administer and collect that cost. Multiobjectives in the model are achieved by the specification of the desired quality to be obtained. Notice that there is no aspect of time in the formulation, that the problem is viewed as steady state under some conservative definition of water body parameters, rather than reflecting the stochastic nature and uncertainty of stream flow, waste flows and physical parameters. Usually, the stochastic effects are considered using the time varying descriptive models discussed in the previous section to evaluate the treatment configuration derived from management model studies.

The decision variables x_j represent the amount of present waste discharge that will be removed and are specified to always be non-negative. Thus, no source is allowed to reduce its treatment level and there can be no further degradation of quality conditions. Where present quality levels are high, they will remain high. This would preclude an attempt to bring in new industry to an area to make use of assimilative capacity available above that which it required. The concept of new polluters at a future time and how they would receive allocations is another policy question that would have to be considered.

Attempts at solving the problem are aided by a linear assumption for the constraints and by the form of $f_j(x_j)$, which is assumed convex and approximated by linear segments. Sobel (94) and later ReVelle et al., (85, 86) Loucks et. al., (66) presented linear programming solutions using linear approximations of the convex cost functions. Liebman and Lynn (61) proposed a dynamic programming model for the non-estuary case.

Attempts at bringing more equity to the solution have also been attempted. One such formulation is the "uniform treatment" model which requires that all polluters treat their waste loads to the same percentage removal while meeting required standards as before. This model can be written as:

$$\text{Minimize } S \tag{A6}$$

subject to:

$$\sum_j B_{ij}x_j \geq b_i \quad i = 1, \ldots, m \tag{A7}$$

$$\left.\begin{aligned} P_j + \frac{x_j(1 - P_j)}{\ell_j} &= S \quad \text{if } S \geq P_j \\ x_j &= 0 \quad \text{if } S < P_j \end{aligned}\right\} \tag{A8}$$

$$0 \leq x_j \leq \ell_j \quad j = 1, \ldots, n \tag{A9}$$

where

B_{ij} = the unit response of water quality in section i due to a unit change in load at source j;

S = the uniform treatment level for the region defined as percent removal of the initial waste u_j at each source;

P_j = the present percent of waste u_j removed;

ℓ_j = the amount of waste presently being discharged at source j after P_j percent removal of the original waste loads;

x_j = the amount of the presently discharged waste at j to be removed;

and the other symbols are as previously defined. Note that $\ell_j = (1 - P_j)u_j$ for each source.

This formulation presents no problems if no treatment facilities already exist in which case the total waste discharge ℓ_j from each source is u_j. But if some treatment already exists, a constraint such as Equation (A8) is needed. That allows the choice of a uniform treatment level below the existing treatment if it is acceptable. While the problem of solving Equations (A6) through (A9) would appear quite difficult as there are discontinuities in the constraint set, in fact optimal solution is obtained by a simple iterative process in which S is increased incrementally until the goals are met.

This uniform model is essentially the same as the efficiency model in Equations (A3) through (A5) but has an added constraint, i.e., Eq. (A8) which means that the regional treatment cost for the uniform model must be at least as great as that for the efficiency model. The uniform model's "equity" feature of allocating costs to users more fairly by requiring all users to provide treatment to the same specified treatment level is the reason for this increase in regional cost. However, closer examination of the features of the plan reveals it has its faults. Requiring the same percentage of waste reduction ignores the fact that the size of waste discharges and the costs associated with fixed improvements vary greatly from source to source. Thus a high degree of removal at a site of a small waste source contributes little to quality but does increase costs at that source. Similarly, the use of a uniform treatment level requires that all sources including those having no effect because of their location must raise their treatment level to obtain increased quality at a particular point. However, the model does have the appeal that it is easy to administer in the sense that each polluter is told directly the percentage of waste he must remove and no attention need be given to cost allocation as it has to be internalized.

A compromise between these two extremes has been suggested. It is the zoned uniform model and requires that sources be divided into zones which reflect equivalent activity, location and/or damages created and that uniform treatment levels be found for each of these zones so that a given quality goal is reached for the region at least cost. This model is formulated as:

$$\text{Minimize} \quad \sum_{j=1}^{n} f_j(x_j)$$

subject to:

$$\sum_j B_{ij}x_j \geq b_i \quad i = 1, \ldots, m \qquad (A10)$$

$$\left.\begin{array}{r} P_i + \dfrac{x_j(1 - P_i)}{\ell_j} = S_k \\ \text{if } S_k \geq P_j \\ x_j = 0 \\ \text{if } S_k < P_j \end{array}\right\} \begin{array}{l} k = 1, \ldots, K \\ j \epsilon R_k \end{array} \qquad (A11)$$

and

$$0 \leq x_j \leq \ell_j \quad j = 1, \ldots, n \qquad (A12)$$

where

S_k = the uniform treatment level in zone k;
R_k = the set of sources j in zone k;
K = number of zones;

and all other symbols are as defined previously.

An optimal solution method based on a branch and bound method has been suggested for this problem by Liebman and Marks (63). Obviously, the solution obtained depends greatly on the way the waste sources have been assigned to zones. If each zone were an individual source, the problem becomes the least cost model, while with only one zone the problem becomes the uniform cost model.

While the previous models considered modification of an effluent at its individual source, there are apparent economies of scale to be gained by building regional facilities and transporting wastes to them. There is a more difficult problem than the previous ones as the range of decision variables has been increased from at source treatment only to include the building of piping systems, relocating waste discharges and the allowance for siting a few large-scale, regional-treatment plants. Such an increase in decision variables causes serious non-linearities in the cost function and in predicting the cause and effect relationships of pollutional discharges into water bodies, particularly when the shifting of a load means a change in the hydrodynamic flows in the stream. Though several attempts have been made to deal with this problem there have been severe computational problems. Graves et al., (36) first studies a problem which allowed only piping of the wastes to areas of better quality, but did not include treatment as a variable. This caused problems because it allows degradation of areas of better quality in order to improve areas of extremely bad quality. It was in effect a redistribution of wastes to make use of available assimilative capacity. In a later work, which is available through the EPA, Graves et al., (37) proposed another model which allows both treatment and piping to be decision variables. The objective function is to minimize the cost of treatment plus the cost of building transportation pipeline while meeting requirements for improved stream quality at specific points. The regional treatment facility problem will continue to be one of great modeling interest due to recent legislative impetus and apparent economies.

Flow augmentation is a control alternative which attempts to increase the assimilative capacity of the stream by increasing the amount of dilution water present during low flow periods. Such a method is quite expensive as it involves the allocation of upstream storage in reservoirs to this use and is to be considered only after adequate waste treatment has been applied. Grantham et al., (34) in work sponsored by the EPA, employ an economic description of the value of low flow augmentation by constructing an optimization model of least cost effluent control which shows how the cost of regional treatment decreases with increased flow in the system. Thus the effect of the flow augmentation is developed not in the objective function but in the constraint coefficients which describe the cause and effect relationship in the physical system. Young and Gitto (111) report on using modeling of flow releases to control acid discharges in streams.

Another method for increasing the assimilative capacity of water bodies is to increase the oxygen supply by artificial aeration using large-scale mechanical equipment. This is an alternative only justifiable in extremely polluted rivers where treatment related schemes have been applied but cannot provide enough improvement. Work in this area to size and space the equipment is presented in Tarassov et al., (98). The optimization technique used is Pontryagin's Minimum Principle and represents one of the few attempts at applying control theory in this area. Ortolano and Thomas (78) suggest simple models for evaluating the removal of bottom deposits.

As noted earlier, a major pollutional load to water bodies is urban storm water. A descriptive model, the EPA Storm Water Model, has previously been discussed which for a given configuration of control, storage and piping facilities will show the magnitude and quality at any point in the system in response to a design storm (60). The problem with a descriptive model, such as this, is that it is relatively expensive to do sensitivity analysis to compare many alternative designs. Optimization serves as a possibility for building a screening model that will suggest good alternatives for the simulation process. Kirshen and Marks (57) report on a linear programming model in which storage location, storage size, additional piping and treatment, location and operation are decision variables. The objective function is to minimize construction and operation costs while meeting treatment and flooding constraints. They report good success in developing meaningful design configurations for further consideration using the simulation model.

The previous models have all dealt with the management of wastes in a regional context from the viewpoint of the technologic controls to be applied. As suggested by Kneese and Bower (58) another approach is to look for pricing schemes, i.e. charges or taxes, which will cause polluters to internalize pollution costs and thus shift their activities. The main advantage for such a scheme is that it allows each waste source to make its own optimal adjustment rather than having a method externally prescribed. This leads to a degree of decentralization which allows for more rational individual decisions. Several researchers have investigated the problem of finding a set of prices for a region which would bring about desired water quality levels. Haas (45), Haimes (39) and Upton (106, 107) discuss multilevel optimization techniques for the problem which characterize the different levels of decision (the choice of quality levels, the setting of prices, and the user response). The main difficulty in such work is not in the premise of such charge schemes, but in their acceptance and implementation. There has been strong reaction against pricing schemes in principle and little attempt by the government to develop a large-scale, computational example to evaluate such a method.

It is clear that future population and development pressures will cause even stronger interactions between pollution, the recovery of materials and the more careful use of resources. In this case, pricing schemes for the resource system as a whole seems an area of important future modeling and consideration. Dorfman and Jacoby (25) examine regional policy by identifying different interest groups within a region and showing how weighting their preferences develop different quality levels and public expenditures in a region.

Energy Production and Thermal Pollution. Recent models to consider new patterns for the expansion of electrical energy production have explicitly considered environmental standards as a factor in site and technology choice. Farrar et al., (29) report on a large scale linear programming model to investigate the supply for a given energy demand pattern in a region at least cost of construction and operation, while meeting reliability and environmental constraints. The environmental constraints are considered by evaluating each site available in a region for the types of generation and abatement technology that might be located there and still meet given temperature and air pollution standards. Shiers and Marks (92) report on a water pollution model which develops this information for a variety of different receiving water and technology alternatives. The result is a series of models which displays good generation expansion schemes that meet environmental standards and which can be used to demonstrate the economic impact of environmental standards on the cost of energy production.

Technology Design. Management models have also been used to design the technology systems for handling and treating wastes. Evenson et al., (27) reports on a dynamic programming model of an industrial waste treatment system and Liebman (62) shows a heuristic algorithm for designing gravity sewer systems. Dynatech (26) has developed models for optimally configuring electric generation equipment.

C. The Delaware Estuary, an Example

The first use of a regional management planning model was by the Federal Water Quality Administration in a 1964–67 study of the Delaware Estuary which runs through New Jersey, Pennsylvania, and Delaware. The problem included 90 major waste sources, approximately 24 quality constraints, and piecewise linear approximations for cost functions and was repeatedly solved using linear programming procedures to evaluate a variety of different assumptions about goals, stream and economic parameters.

In designing water quality improvement programs for the Delaware Estuary, the Federal Water Quality Administration generated several alternative schemes to reach a range of quality goals. As they could demonstrate that the technical alternatives such as low flow augmentation, piping, stream aeration, and storm water control were clearly dominated by selecting waste treatment levels at each source for the particular conditions on the Delaware, the latter alternative was the only one considered. No consideration was given to taxing or incentive schemes or to the possibility of regional treatment.

Six sets of alternative water quality goals were chosen with the help of citizen groups, conservationists, industrialists, and governmental representatives. Objective set I represented the best that could technically be obtained by requiring maximum waste reduction from the 90 largest waste producers using the estuary. At the other extreme, objective set V represented the cost to keep the conditions as they were at the time. As can be seen in Table A.1, a considerable expense is needed just to maintain the status quo. The objective sets II, IIA, III, and IV represent various possible alternatives in between. The goals are noted in Table A.1, along with the cost to meet the goals under different cost allocation assumptions and with some estimates of benefits.

Costs, including capital costs and the present value of operation and maintenance expenditures, were estimated to obtain each quality objective set using the least best, uniform, and zoned uniform treatment models. The quality to be obtained from each objective set included the valuation of visual, esthetic, and public health benefits, as well as the quantifying in dollars of fisheries, recreation, and water supply. These costs and benefits are presented in Table A.1.

There are several important factors to be noted about the costs. First, the cost of maintaining present conditions over the next 15 to 20 years is significant. Without action things will get worse. In the lower quality ranges (objectives II-IV), the difference between a least cost and uniform treatment model is quite significant. To reach higher quality levels, it must be remembered that there are many waste sources over which there is no readily available control. Thus, to account for these, more of the treatable wastes must be removed. Since the cost of treatment versus the percentage of waste at a source removed is highly convex, costs go up quickly. As well, there are relatively few waste sources under control and meeting higher goals requires that almost all be removed, thus giving little difference between allocation models.

For this particular region, an institution exists for the region specifically designed and empowered to deal with water resource problems, the Delaware River Basin Commission (DRBC). Such figures as those in Table A.1 pro-

Table A.1

Capital and Present Value of Operation and Maintenance Costs in Million Dollars to Obtain Stated Objectives under Three Types of Management Models (Dollar value of some of the benefits for each objective are given.)

Objective Set	Least Cost Model	Zoned Uniform Model	Uniform Model	Evaluation of Some Benefits ($Million)
I	695	695	695	492
II	442	493	509	430
IIA	387	430	439	430
III	291	326	383	303
IV	238	256	319	282
V	140	140	140	162

vide excellent supplementary information for such a decision-making planning body to use along with its basic political, social and economic knowledge of the region. The plan finally chosen by the DRBC was a compromise between two originally specified. Objective set III was strongly championed by industrial and governmental interests who had a large economic stake in the outcome. Objective set II or I was the choice of the various conservation and civic groups. A compromise, objective set IIA, was selected and the DRBC is now involved in the problems of implementing, financing and overseeing the implementation process.

D. Summary of Water Quality Models

From the above discussions, we see that both the major categories of descriptive and management models are used in a wide range of decision making problems. A proper evaluation of the impact of such models in the total water quality decision framework must, of course, lead to a conservative conclusion. Much needs to be accomplished to bring the researcher, with his concerns about model structure and data and grinding out a solution, closer to the decision maker who is constrained by a diversity of forces which range from budget to technological to specific environmental considerations. In most cases, institutional progress has been made and river basin authorities have been set up within a legal framework that makes implementation possible. But even a regional authority must still decide how it will assess costs and make large-scale decisions. More attention must be paid to modeling of the implementation process and the political market place in which the various objectives and interest groups negotiate solutions. The existing management models have addressed problems which, while computationally difficult, were in fact easy to conceptualize. The fact that there has been difficulty in implementing results stems naturally from two problems. In some cases researchers have addressed themselves to problems that interested them, not the people who had to make the decision. In other cases, the modeling process was seen by the analyst as synonymous with the decision process. This is just not possible for there is no way to capture the complexity of the decision process in a model and the analyst must be content to provide information which helps illuminate for the decision maker the tradeoffs available in the decision process.

It is clear from the consideration of descriptive models that only a small beginning has been obtained in the vitally important questions of translating physical system responses to biologic and ecologic responses in the surrounding habitat and to economic, social and demographic responses in the surrounding region. These are very difficult modeling questions, not only because such processes are suspected to be highly dynamic and non-linear, but because there is great difficulty in conceptualizing their structure.

An important trend is the realization that descriptive models and optimization models are not competitive model structures but processes to be used jointly to attack problems. Each can provide information and guidance in the area where the other is weakest. More attention will be placed in the future on applying these techniques to modeling in systems not yet fully understood such as ecologic and economic systems. Also important is the consideration of waste management and resource use as an integrated system. Further investigation must be given to land use-water quality interactions and to policies that prevent pollution from reaching receiving waters such as zoning, better construction practices, etc. Pricing schemes in resource use and pollution control deserve much more attention than they have received for they provide a rational structure for coordinating a whole series of interactive yet seemingly diffuse systems. Thus, considering pricing schemes really becomes an attempt to expand the system's boundaries to get at some of the exogenous effects which are driving the system.

II. WATER QUANTITY

A. Description of the Problem

1. Problem Identification

The general problem of water resources is the allocation of a resource that is distributed in space and time in the proper quantity and quality to a set of often conflicting alternate uses. In the previous section, we dealt with that aspect of the problem that relates to water quality and the way pollution control may be evaluated in the context of having water available in sufficient quality to allow desired water uses. In this section we deal with the quantity aspects of water. This involves the evaluation and decision processes necessary to distribute water differently than it is distributed now. In some cases the problem of water quantity is that there is too much on a particular place as in the cases of flooding, drainage and water logging or agricultural lands. In other cases, it is supplying more water at a particular location for a particular use than is available there. Both cases may involve the building and operation of large capital intensive structures for storage, transportation and treatment to change the physical phenomena that do occur. To arrive at the correct distribution of water among its various uses is a problem in public decision making for much the same reasons that water pollution is a public problem. The normal private market system, based on private ownership and price for exchange as a measure of preference, breaks down in the face of public resources such as water, particularly when the scale of investment necessary to reap the benefits of water development is so large. In the absence of a market mechanism, the decision process must make allocations among various water uses based on some form of analysis which reflects the multiple objectives for its use and the preferences for different ordering of objectives by different interest groups. Consider a reservoir built to store water for several different purposes: recreation, water supply, flood control, power generation and flow releases for downstream uses, such as agriculture or low flow augmentation. Inflow to the reservoir is a stochastic variable with high trends in the spring and low trends in the late summer and early autumn. For flood control purposes, the reservoir is best operated by leaving it empty so that it has capacity to dampen flood peaks should floods occur. For uses such as irrigation, it is desirable to fill the reservoir in the spring so that the water can be released when it is needed in the summer. For uses such as recreation which depend on a full reservoir, it is best to store water and to keep it stored. Thus, there is a conflict in that allocation or operation of the system for one use means less is available for another use. It is in the following areas in which the model builder can provide valuable information: (1) establishing what is expected to occur in the system due to natural processes, (2) how control alternatives and allocation to different uses impact on these processes, (3) what benefits are gained by a particular use, and (4) what are the pertinent tradeoffs between objectives.

The basic framework for the consideration of the multi-objective aspects of water allocation is presented by two major works from the Harvard Water Resources Program in the early 1960's (see Maass et al., (68), and Marglin (73)). They helped to demonstrate the need for the consideration of objectives other than the traditional efficiency or national income measures which had been and, in some cases, continue to be the single objectives for project evaluation. Other objectives for consideration which vary from project to project also include regional income, environmental quality, national defense and other possibilities. The problems in considering multiobjectives is two-fold. First, suitable measures of effectiveness must be chosen for each objective that reflect the major concern for it. Such measures need not necessarily be in commensurate units: national income may be in dollars, while environmental quality may be measured in miles of scenic river preserved. Second, accounting systems must be set up which show the impact of a particular project configuration on the relevant objectives. This results in a net benefit transformation surface such as shown in two dimensions in Figure 1, in which all feasible solutions are ranked according to their impact on the defined objectives and the envelope of the surface represents a superior set of feasible alternatives in the Pareto sense. That is that every point on the surface is feasible and exhibits the property that any move along the surface implies gain for one objective only at the expense of loss for another objective. Choice of the proper place to be on the surface is a function of the social welfare function of society as shown in the hypothetical surface and indifference curve intersection at point A in Figure 1. The difficulty comes not only in developing what the net benefit transformation surface looks like but in determining what the shape of the social welfare function looks like. It is well demonstrated by Arrow (3) that such a gross measure of society's desires can not be found by aggregating individual or interest group preferences even if they could be assessed. Thus, the role of the analyst becomes that of showing what the transformation surfaces between different objectives look like and perhaps hypothesizing the preferences of different interest groups. Here the role of models for the water resource decision process is so prevalent throughout all phases of the investigation as to make it difficult to distinguish model building in water resources from water resources analysis itself. Models are addressed to understanding and describing the underlying processes that transform rainfall into surface runoff and groundwater, to determining what are the effects of building control and transportation technology, to predicting the impact of water development on the surrounding social, economic and political environment, to decide on when and what to build and the best way of operating a system. The next part of this section deals with the different types of water use, how allocation of quantities to one use can conflict with other uses, and to what system objectives each use contributes. Then the various aspects of water resource analysis are discussed by functional form of the specialists who deal with individual problems. In section B the role that models play in developing information both from the descriptive and management sense are presented. Then a case study for the Rio Colorado, Argentina, developed by the MIT Water Resources group is briefly presented to show what types of models were used at the various decision levels of the problem.

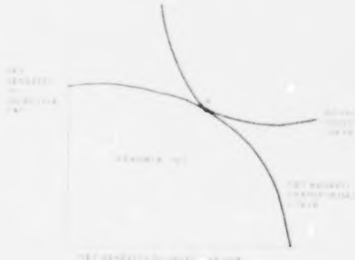


FIGURE 1—*Multi-objective Analysis*

2. *Water Use*

Wolman (108) in his article "The Metabolism of Cities" describes the uses to which water is put to in the U.S. These uses are indicative of the quantities involved in present consumption and of preferences in future consumption. What is striking is the way the uses are presently distributed and the question of how future demands will be met. Of the 4200 billion gallons of water per day of precipitation that falls each day in the U.S., about 40% is utilized where it falls for the growth of vegetation, forests and crops, while another 30% evaporates directly from the soil or returns to the atmosphere through evaporation from vegetation. Of the remaining 1200 billion gallons available for other uses, about 800 billion gallons are reasonably available in terms of transportation, institutional and quality problems. Major uses include:

a. *Irrigation of Farm Lands.* This represents about the largest usage of water in the order of about 140 billion gallons per day in 1960. Since irrigation implies supplying water where it is not normally available, this means that water from somewhere else must be diverted to the irrigation area and away from other potential uses. Further, the use of water for most irrigation uses depletes it both in quantity through losses and in quality through the chlorides and nutrients picked up in the irrigation process. Traditionally, the use of water for irrigation has been seen as having both strong national income and regional development components, while the return water implies an environmental problem. To this end, large public investments have been made to transport these large volumes of water often over great distances so that irrigation can take place. The price of such water, either actual or implied, is far below the price charged for other uses.

b. *Industrial Use.* Industries withdraw water for many purposes ranging from process water to cooling water and waste dilution flows. This use in 1960 in the U.S. was on the order of 60 billion gallons per day and the return flows are often diminished in quality and quantity. Objectives here are both national and regional income, as well as a negative impact on environment.

c. *Cooling Water for Steam Electric Utilities.* Electric power production using steam driven turbines requires tremendous volumes of cooling water which are normally returned without consumption but with increased heat content. This use in the U.S. in 1960 was about 98 billion gallons per day. Projected future cooling water needs as electrical energy demands increase are enormous. Relevant objectives include national and regional income as well as the environment.

d. *Municipal Use.* Municipal use of water normally averages about 150 gallons per person per day for washing, consumption and waste. While this seems like a great deal of water, the total usage of about 23 billion gallons per day is only about 8% of the total water usage. Here water is used for development, and aesthetics and return water causes environmental problems. Other important water uses do not necessarily require the withdrawal of water from its natural source, but may require additional flows to allow use to take place there. These include navigation, water based recreation, aesthetics, power production, support of fish and wildlife, commercial fisheries and waste disposal.

In a more subtle and less obvious fashion, water can also be seen as a necessary prerequisite and principal focus for development of other activities. Water resource investment by government is often used by governments not only to increase income for the nation but to redistribute it among the citizenry. Often projects such as TVA have a strong regional development focus. Allocation of water to various uses have strong implications of our priorities for development and the necessary ingredients such as power and agricultural goods for that process. There are an enormous number of governmental agencies in the water resources business with the main construction authority going to the U.S. Army Corps of Engineers (Defense), the U.S. Bureau of Reclamation (Interior), and the Soil Conservation Service (Agriculture). Each exists mainly for different objectives, yet each builds multipurpose developments. Given the future stress caused by more and more demand for a fixed resource, it is clear that water will be reused more efficiently and that projects will seriously interact. Thus, careful attention must be brought to the allocation process. The decisions to be made are either at a governmental level or, if private, with the approval of the government. In most cases, governmental units such as the Corps of Engineers design and build their own works. In local government, often both the design and construction is subcontracted.

3. *Functional Groupings for Problem Solutions*

As is demonstrated by the complex technical, social, political and economic interactions in water resource development, many different talents are necessary in the analysis of water quantity problems. Several fields of subinterest can be identified and, of course, there is major overlap in many areas. These categories are discussed next.

The field of hydrology and hydrologic engineering is concerned with the understanding and analysis of the rainfall runoff relationship. Interest centers about the statistical properties of rainfall and flood events in order to predict and simulate their occurrence, the modeling of catchment response to predict the runoff quantities and timing for rainstorms and the design of structures and land use controls for controlling floods and providing water for other uses. A major subfield is *ground water hydrology* which deals with questions of water movement and availability underground. Because of the highly stochastic nature of rainfall amounts and timing, there is a strong statistical basis to this work. Further, urbanization of catchments leads to changes in the quantity and quality of runoff. The subfield of *urban water management* deals with the particular effects of urbanization and development on water resources. Normally, the hydrologists work both in government and in private practice to predict ahead of time the risks involved in building in a particular area, the impact of new construction and the control alternatives available for ameliorating present or future problems.

Hydraulics and Hydraulic Engineering deals with the mechanics of water flow. Problems of interest are the movement of water in channels and other conduits, the structures necessary to create power or provide storage and the operation of systems for multiple use.

The job category of *Water Resource Planner* or *Water Resource Engineer* requires knowledge of the other two areas plus integrating knowledge on the evaluation of projects. This requires an understanding of economics and institutions, as well as of the mathematics of systems analysis for planning system configurations. Often this function is carried out as a team effort with different disciplines handling different parts of the job within the framework of the decision information to be generated. There are two distinct levels of approach which divide along the fields of the descriptive and management models described in the next section. The first field requires the ability to model cause and effect relationships of particular actions. This area is simulation oriented and relies heavily on regression, dynamic modeling and the computer to predict system responses. The second field implies the selection among alternatives of the best configuration. Modeling at this level is both simulation and optimization.

B. Models in the Aid of Water Quantity Decision Making

As mentioned previously, the models to be discussed impact at different levels in the decision process from the broadest scale allocation problems to the microscale of every day operation of a system. A distinction will be made between descriptive models which attempt to show the cause and effect relationship between a particular policy and the important parameters of the system, and management models which attempt to choose among policies.

1. *Descriptive Models*

An important question in hydrology is the simulation of rainfall-runoff events, and modeling is used extensively in this process. One important area is computer modeling through simulation of the actual physical phenomena involved in the process. Crawford (21) describes the Stanford Watershed Model, and Harley (44) describes the MIT Catchment Model. Each of these represent large scale computer based attempts at organizing the important data for a catchment such as degree of imperviousness, amount of vegetation, population and geographic conditions, and predicting the timing and volume of runoff at any point in the system for a given storm. Such models are used routinely in practice for several purposes. First they can be used to simulate proposed control schemes and to evaluate the level of flood security at a site. They can also be used to predict downstream effects due to changes in the catchment. One of the most crucial issues in development is the interaction between changes in land use and water resource characteristics. As the use of land changes from a natural state to one of more intensive structural and economic activity, runoff and flooding problems are accelerated and water quality conditions are affected. What makes the problem even more important is that land use planning rarely takes this into account. The remedial measures for control, once the land development has taken place are costly and difficult because of the dispersed nature of the problem. The quality issue is very important in urban catchments where much of the street litter and air pollution fallout are caught in storm water drainage. The work of Lager et al. (60) and of Chen (18) are directed towards these models. All the models are fairly heavily data dependent but computationally able to represent reasonable sized catchment areas. They are all simulation models in the sense that a model of the physical system is operated mathematically over time in response to a given input event.

Another important issue in hydrology is an attempt to gain statistical information about the time spacing of storm events and of the magnitude of river flow quantities. The ability to statistically generate long time series of data which have similar characteristics to short available data records has been one of the most important aspects of the field. Fiering (31) summarizes the theoretic basis for this synthetic generation of data which has become an accepted part of the design and planning process. Other work in this area include important papers by Mandelbrot (71), (72), Tschannerl (105) and the development of a great many computer routines by the Hydrologic Engineering Center (HEC) of the U.S. Army Corps of Engineers (51). The HEC is instrumental not only in computer model development in hydrologic engineering problems, but also serves as a major teaching and research center for other government agencies involved in this field. Examples of their work include papers by Beard (5), (6), and large scale computer packages for a variety of hydrologic functions (19), (20). Other areas which depend on the knowledge of the statistical character of rainfall and streamflow events are the design of reservoirs for the collection of water for use at a later time and to prevent floods by dampening flood peaks. The flood problem of course has received considerable interest. Day (22) has developed a model for the consideration of investment in non-structural flood control, while Bhavnagri (7) has developed a stochastic model for flood proofing in a flood plain. Askew (4) has considered the problem of the statistical aspects of critical droughts and many of the HEC programs, particularly Beard (5), deal with the statistical definition of safe (guaranteed) yield from a reservoir.

Groundwater presents an important modeling area with much attention being placed on attempting to analytically or through simulation describe the spatial and temporal aspects of its movement and availability. Breitenbach (10) describes many of the digital computing aspects of such modeling. Freeze (32) has developed very complex models of the physical processes in groundwater which are very precise, but have enormous data and computational time requirements. Kleinecke (56) describes the use of linear programming for the establishment of parameters for groundwater systems. Thus, while we most often think of optimization as a tool for choosing among alternatives in an economic environment, in dealing with physical systems, it may also be used to identify important parameters of the system. Pinder (81) and Taylor (99) are further examples of the role of computer models and analysis in the description of groundwater movement.

A most important aspect in the development of water resource planning is not only the estimates of cause and effect relationships in the physical system but in the surrounding social and economic systems as well. It has long been realized that new water resource development can bring about important changes in demographic variables, income redistribution and other factors. Because these aspects are much more difficult to model they are only just beginning to receive attention. On a regional basis, Hamilton et al. (42) use systems dynamics models to show the demographic and economic aspects of developments in the Susquehanna River Basin. Lofting (64, 65) uses input-output models of the economy to show the value of water development in an economic context. Such input-output analysis was also used by Schaake (91) in the estimates of regional demand for water in the future. More detailed studies of residential demand using statistical modeling are reported by Howe and Lineweaver (49) and Hanke (43). Tadros (97) uses both statistics and optimization to estimate recreational demand for water resources projects. Haissman (40) also uses optimization to determine manpower needs for irrigation developments in Mexico.

2. *Management Models*

The consideration of management models which investigate and choose among a series of alternatives requires the resolution of tradeoffs between model reality and computational ability. The trend in the past few years in this field has been from very complex optimization models, which could not be solved for large scale cases, to relatively simpler optimization models, called screening models, which could suggest solutions for large scale problems for further investigation. The philosophy to be evolved here is one of the conjunctive use of optimization and simulation techniques in management models, each technique having complementary strengths and weaknesses.

The basic problem in water resource planning is to decide on the optimal timing and investment in capital structures and operation of a given system to meet stated objectives. The attendant problems in solving or even formulating such a general management problem are many fold and include the stochastic nature of the system both in hydrologic and economic terms, the difficulty in identifying and dealing with multiple objectives, and size and dimensionality. We are thus led to a subdivision of different models, each directed towards an aspect of the general problems. These models are used in the spirit of supplying information and understanding of the decision-makers problems, not in optimally solving them.

We first consider the problem of choosing between multi-objective projects. Here there are conflicting desires by different interest groups for the outputs of the system, and difficulty in developing and comparing benefits to the different objectives. Models directed towards the problem include the following. Major (69) in a paper on the North Atlantic Region Water Resources Study of the U.S. Corps of Engineers tells of the role systems analysis has played in large scale plan formulation for water supply. More theoretical methods are suggested by Marglin (73) and Major (70). Cohon and Marks (16) present computational methods for generating transformation surfaces between multiple objectives using linear and integer programming and identify both objective function and constraint methods for representing the transformation surface. Haith (41) describes models for assessing preferences for different objectives by different groups. Rogers (89) uses game theory and mathematical programming to show the payoff matrix for project development between two nations on an international river when there is the possibility of cooperation in planning.

Management models have a significant role in the actual operation of individual components of the system. In particular, reservoirs are often run for different objectives such as power generation, flood protection, water supply, and recreation which require different modes of operation. The important question here is how much water should be stored and released in the system when the inputs to the system for the time period are stochastic. Roefs (87) summarizes many of the attempts at mathematically modeling such problems. ReVelle (85, 86) in a series of papers on reservoir operation uses chance constrained linear programming and a linear decision rule to design operating policy for a reservoir. Joeres et al. (54) consider the operation of an interconnected system using both linear programming and a stochastic simulation. Loucks (66) deals with the concept of using Markov models to determine optimal reservoir operating rules, and Young presents a dynamic programming model (109). Bodin (8) and Butsch (15) have also done work in this area. Several authors have addressed the issue of using mathematical programming in the design of multi-component systems. Jacoby and Loucks (52) deal with optimization/simulation interactions on a large scale case study for the Delaware system. Hendari (46) and Butcher (13) also present programming models for this type of investigation. The problem of optimally determining where given water supply demands will be served from is dealt with by deLucia (23) using a linear programming model for the North Atlantic region. Demand estimates are derived from the earlier described works by Schaake (91), and the optimization problem was to match up supply and demand while minimizing cost to the region. Another important issue in management models is the issue of the time that investment should be made. Such considerations are brought about both by budget constraints which limit the resources available for construction in any one time period and by the existence of benefits for projects that change over time as a result of exogenous market changes. Because of the extra dimensionality added by the inclusion of time, such considerations are usually addressed in separate models. The easiest consideration is the capacity expansion model which assumes that there is no interaction between projects and that demands for services are increasing. The numerous literature in this area, mostly based on dynamic programming is represented by Morin and Esogbue (76), Scarato (90), and Butcher, (14). In cases where there is more interdependence, dynamic programming proves more difficult and other types of programming models are necessary. Young (110) and Facet and Marks (28) both use linear and integer programming algorithms.

In the Facet model, both a scheduling and sequencing model is presented for highly interactive systems. A scheduling model allows the choice both of project size and construction time as decision variables. Such models are severely limited by number of time periods and decision variables if satisfactory computational results are desired. A sequencing model assumes that a size has been chosen for each project, usually by a screening model plus simulation, and then uses integer programming to decide on an optimal sequence for the projects. While more restrictive in formulation, computation is much easier and much less costly, thus allowing more sensitivity analysis.

The design of the hydraulic systems for transporting water and for processing it for different uses have also been the subject of management models. Buras (11) is an example of the use of optimization in the design of aqueduct size and pumping cost using dynamic programming in a tree type network. Because there is no redundancy in the system, the flow in any link is uniquely known and even though costs are nonlinear, the resulting optimization problem is quite tractable. Local distribution systems purposely have a great deal of redundancy built into them to allow for other system objectives such as reliability. In this case the problem becomes highly interactive because the flow distribution in the system, instead of being fixed, is a function of the decision variables. This more difficult problem has been the subject of a considerable amount of study and work such as de Neufville et al. (25), Kalley (55), and Kohlhaus (59) are indicative of management tools available for aid in distribution system design. Huang (50) reports on the use of optimization in water treatment design.

Agricultural management of water has received attention usually in the context of the use of groundwater and in the conjunctive use of ground and surface water for the purpose of the supply of the large volumes of water necessary for agricultural uses. Rogers and Smith (88) and deLucia (23) describe stochastic management models for the aid of agricultural decision making. Aron and Scott (2) is an example of a large volume of literature dedicated to arriving at the optimal mix of ground water and surface water use. Burt (12) has dealt extensively with the optimal development of groundwater to meet expending needs for agricultural use.

Examples of other management models developed include the work represented by Matalas (74) in the optimal location of gauging stations for the acquisition of data for the design of water resources systems. Grayman and Eagleson (38) report on similar work in the design of radar systems for gathering information for the forecasting of storms.

3. *Case Study—Rio Colorado, Argentina.*

This case study shows a typical water resource problem and the role that models can play in aiding decision-making for resource allocation for the problem. More detailed information can be found in Grayman et al. (37). The Rio Colorado, shown in Figure 2, rises in the Andes Mountains as a result of snow melt runoff and flows 550 miles in an easterly direction to the Atlantic without appreciable inflows from precipitation. The average flow is 5000 cfs and it flows through a very sparcely settled land (44,000 people in 26,700 square miles). The main purposes for management and development of the river are irrigation and power production, with flood control and water supply also being relevant. For these purposes, several major purpose projects have been proposed by Federal and Provincial groups including diversions to the Rio Colorado from the nearby Rio Negro and from the Colorado to other basins. These are shown schematically in Figure 3. The models used to develop plans included a deterministic linear programming screening model, simulation models for showing the response of the physical system and of the demographic sector to water resource investment in the area and an integer programming sequencing model. The screening model was of the form:

Maximize Net Benefits to Each Objective
Subject to Constraints on
  Physical continuity
  Technical possibilities
  Policy Constraints

The means for handling the multi-objective nature of the objective function included a weighting method which involved choosing different weights for the different objectives and a constraint method where minimum requirements were set for one objective while maximizing the other. The objective function itself was assumed to be piecewise linear, with a few integer variables related to whether certain large storage projects should be built. The typical size of such a model when developed for three seasons and one yearly time period (i.e., a deterministic assumption), was on the order of 700 constraints and 700 variables.

An example of the type of information that can be developed on multi-objective decisions from such a model is shown in Figure 4, which shows a transformation surface between national

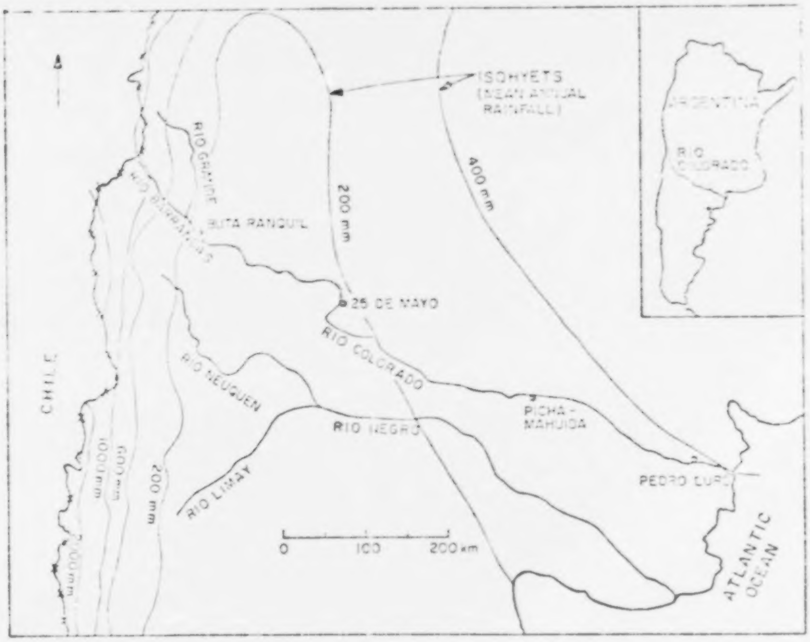

ISOHYETS (MEAN ANNUAL RAINFALL)
ARGENTINA
RIO COLORADO
RIO GRANDE
RIO BARRANCAS
BUTA RANQUIL
200 mm
400 mm
25 DE MAYO
RIO COLORADO
RIO NEUQUEN
RIO NEGRO
PICHA-MAHUIDA
PEDRO LURO
RIO LIMAY
CHILE
1000 mm
600 mm
200 mm
2000 mm
0
100
200 km
ATLANTIC OCEAN

FIGURE 2—*Location of the Rio Colorado, Argentina*

income and some measure of regional equity. National income is measured as net efficiency benefits in monetary units. The choice of a metric for regional equity poses more definitional problems. Here a measure of the total deviation from some equitable distribution is chosen, with the efficient solution in terms of national income having high deviations in water allocation. As a more equitable distribution is sought, i.e. as there is a smaller deviation from the equitable distribution, national income decreases. The point H represents a break point between export and non-export from the basin, i.e. to gain equity in water distribution, water must be diverted from profitable national use, which requires export from the region, to uses within the region. The use of such simple models make it possible to do a great deal of sensitivity analysis and suggest that a problem can be successfully decomposed and that more than one model can be useful in developing information.

Screening models such as these develop sets of configurations that appear suitable for further investigation, particularly to capture the stochastic nature of the system which was missed by the deterministic screening model. In this case two different simulation models were specified for this purpose. One goal was to devise a simulation model that could, with very little input data, provide insight into system behavior. In areas such as the one under consideration, where little development has taken place, the availability of data is often minimal. Thus, a detailed simulation model was built to explain physical processes that were much too detailed for general purpose system simulation. Instead, such detailed simulation models are used to generate input data and parameters for less deailed and thus more operable simulation models. This model uses stochastic data generation by sequentially producing annual values of unregulated stream flow using a multivariate Markov model. These annual

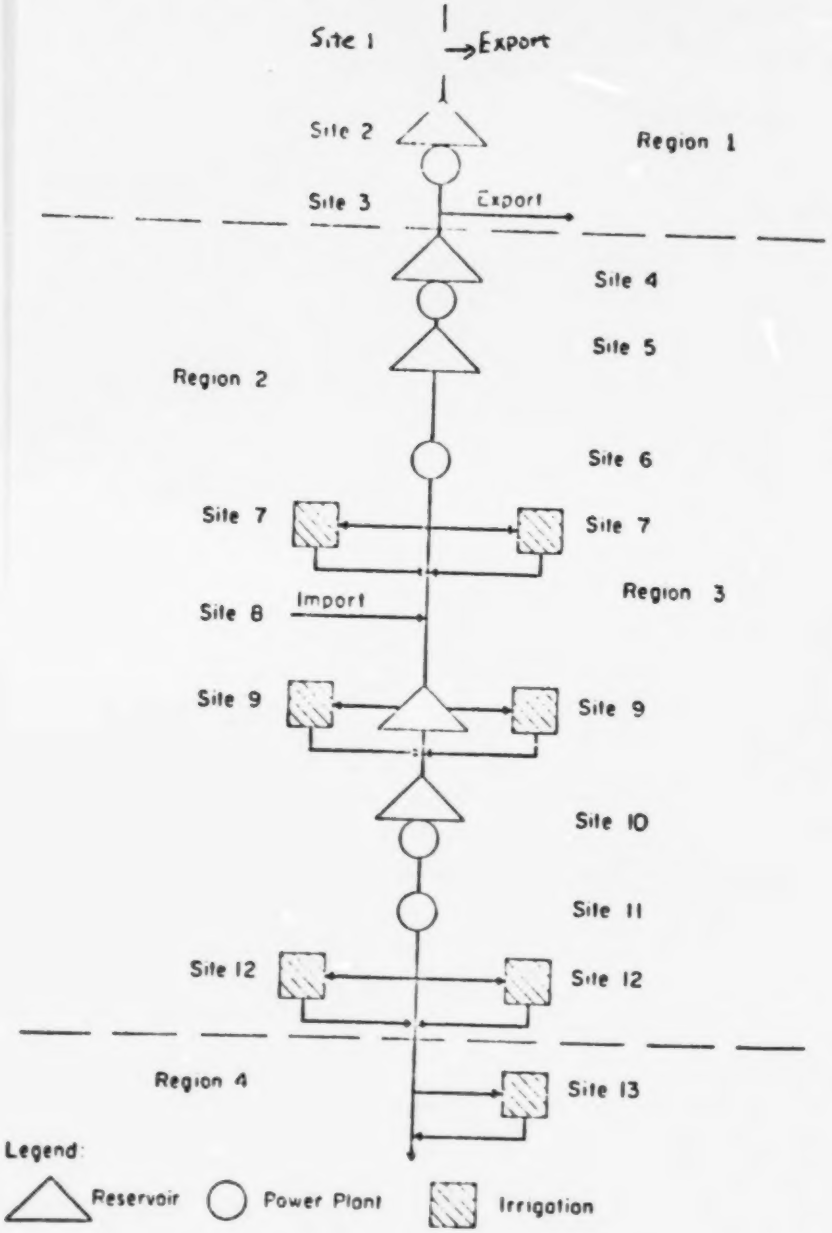

Site 1
Export
Site 2
Region 1
Site 3
Export
Site 4
Site 5
Region 2
Site 6
Site 7
Site 7
Region 3
Site 8
Import
Site 9
Site 9
Site 10
Site 11
Site 12
Site 12
Region 4
Site 13
Legend:
Reservoir
Power Plant
Irrigation

FIGURE 3—*The River Basin and Development Alternatives*

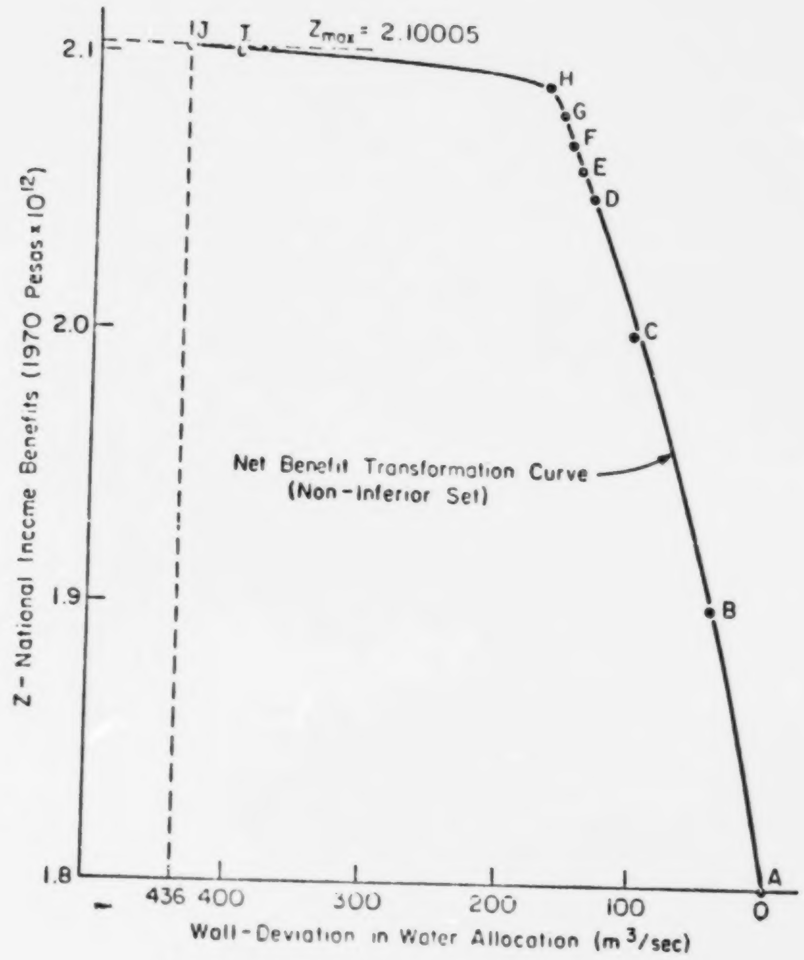


FIGURE 4—*The Generated Net Benefit Transformation Curve*

values are then disaggregated to produce seasonal, monthly or daily values as required, while maintaining the internal means and covariances within each year and between sites. The nonphysical system was also of concern since the demographic effects of this sort of water resource development is vitally important in determining whether to build it. The benefits of the projects will only be reaped if people come to this isolated region and work in agriculture and related industries. To this extent a dynamic policy model based on systems dynamics was developed which showed the feedback relationships and time delays involved in migration and regional attractiveness.

The interactive use of these models, plus additional data and benefit estimates for input to the analysis process, has brought about a series of alternative plans now under review by the government of Argentina. By outlining alternatives and their impacts on different system objectives, the model approach has helped to start a constructive planning dialogue between various regional interest groups in Argentina and, hopefully, the choice of a good consensus plan acceptable to interest groups—a necessity for successful implementation.

4. *Summary*

In the previous sections, we have described the modeling efforts in water quantity. As with water quality, the main emphasis in modeling has been on the development of cause and effect relationships for physical systems and for optimal management. The areas where much more interest in modeling is necessary is in understanding the role of water and changes in its availability on the social and political system of a river basin. First attempts at modeling such phenomena have been mentioned, but this type of work obviously needs to be better understood. Another area where more work is needed is in dealing with the multi-objective nature of the decision-making process. Better means for assessing utility functions for interest groups, for displaying alternatives and for recognizing points at which some agreement can be reached is important from the viewpoint of implementation. In terms of management modeling, the most important future developments will be in better modeling the alternatives available for the control of water resource impacts caused by changes in land use and development policies.

References

1. R. L. Anderson and A. Maass, "A Simulation of Irrigation Systems—The Effect of Water Supply and Operating Rules on Production and Income on Irrigated Farms", Economic Research Service, Fort Collins, Colorado, Tech. Bulletin No. 1431, January, 1971.
2. G. Aron, and V. H. Scott, "Dynamic Programming for Conjunctive Water Use", J. of the Hydraulics Division, ASCE, Vol. 97, HY5, May, 1971.
3. K. Arrow, "Social Choice and Industrial Values", J. Wiley & Sons., New York, 197[illegible].
4. A. J. Askew, W. W. Yeh and W. A. Hull, "A Comparative Study of Critical Drought Simulation", Water Resources Research, Vol. 7, No. 1, pp. 52–62, February, 1971.
5. L. R. Beard, "Methods for Determination of Safe Yield and Compensation Water from Storage Reservoirs", The Hydrologic Engineering Center, Technical Paper No. 3, Corps of Engineering, U.S. Army, August 16, 1965.
6. L. R. Beard, "Simulation of Daily Streamflow", The Hydrologic Engineering Center, "Technical Paper No. 6, Corps of Engineers, U.S. Army, 1967 b.
7. V. S. Bhavnagri, and G. Bugliarello, Flood Proofing in a Flood Plain: A Stochastic Model", J. of Hydraulics Division, ASCE, Vol. 92, No. HY4, pp. 63–76, July, 1966.
8. L. D. Bodin, and T. Roefs, "Determination of the Operation Policy of a General Reservoir System", J. Network Analysis, January, 1972.
9. J. Bradehaeft, and G. Pinder, "Mass Transport in Flowing Groundwater", Water Resources Research, Vol. 9, No. 1, February, 1973.
10. E. A. Breitenbach, "Digital Simulation of Groundwater Systems Simulation of Water Resources Systems", Nebraska Water Resources Research Institute, Lincoln, Nebraska, July 18–23, 1971.
11. N. Buras, and Z. Schweig, "Aqueduct Route Optimization by Dynamic Programming", J. of Hydraulics Division, ASCE, Vol. 95, No. HY5, pp. 1615–1631, September, 1969.
12. O. Burt, "Groundwater Storage Control under Institutional Restrictions", WRR(6)6, December, 1970.
13. W. S. Butcher, "Stochastic Dynamic Programming for Optimum Reservoir Operation", Water Resources Bulletin, Vol. 7, February, 1971.
14. W. S. Butcher, Y. Haimes and W. Hall, "Dynamic Programming for the Optimal Sequencing of W. S. Projects", WRR, Vol. 5, No. 6, December, 1969.
15. R. J. Butsch, "Reservoir System Design Optimization", J. of the Hydraulics Division, ASCE, Vol. 96, No. HY1, January, 1970.
16. J. Cohon and D. Marks, "Multiobjective Screening Models and Water Resource Investment", WRR, to be published, June, 1973.
17. J. Cohon, T. Facet, A. Haan, and D. Marks, "Mathematical Programming in River Basin Planning, M.I.T. Technical Report, Department of Civil Engineering, January, 1973.
18. C. Chen and R. Shubinski, "Computer Simulation of Urban Storm Water Runoff", Proc., J. of the Hydraulics Division, ASCE, Vol. 97, HY2, February, 1971.

19. Corps of Engineers, "Computer Applications in Hydrology", Proceedings of Seminar on Computer Applications in Hydrology, Corps of Engineers, Hydrologic Engineering Center, February, 1971, Davis, California.

20. Corps of Engineers, HEC 3, "Reservoir Systems Analysis Computer Program 723–X6–L2030", Hydrologic Engineering Center, Corps of Engineers, U.S. Army, Davis, California, February, 1971.

21. N. H. Crawford and R. K. Linsley, "Digital Simulation in Hydrology: Stanford Watershed Model IV", Technical Report No. 39, Stanford University, Stanford, California, July, 1966.

22. J. C. Day, "A Recursive Programming Model for Non-Structural Flood Damage Control", Water Resources Research, Vol. No. 5, pp. 1262–1271, October, 1970.

23. R. J. deLucia, "The North Atlantic Regional Supply Model", 52nd American Geophysical Union Meeting, Washington, April, 1971, WRR, 1972.

24. R. de Neufville, J. Schaake, Jr., and J. H. Staflod, "Systems Analysis of Water Distribution Networks", JSED, ASCE, Vol. 97, SA 6, December, 1971.

25. R. Dorfman, and N. D. Jacoby, "A Model of Public Decision Illustrated by a Water Pollution Policy Problem", The Analysis and Evaluation of Public Expenditures: The PPB System, Joint Report, Subcommittee on Economy in Government, 91st Congress, 1st Session, May, 1969.

26. Dynatech Inc., " A Survey of Alternate Methods for Cooling Condenser Discharge Water—System Selection Opt.", WPCR Series No. 16130, DHS, Environmental Protection Agency, Washington, D.C., January, 1971.

27. D. E. Evenson, G. T. Orlob, and J. R. Monser, "Preliminary Design of Waste Treatment Systems", Industrial Water Engineering, Vol. 16, No. 1, pp. 16–21, February, 1969.

28. T. Facet and D. Marks, "Scheduling and Sequencing in Water Resource Investment Models", Inter. Symp. on Water Resource Planning, Mexico, December, 1972.

29. D. Farrar, F. Woodruff, Jr., D. H. Marks and F. Schweepe, "A Model for the Determination of Optimal Electric System Generation Expansion Patterns", Report to the National Science Foundation, M.I.T. Energy Lab., Cambridge, Mass., February, 1973.

30. K. D. Feigner and H. S. Harris, Documentation Report, FWQA, Dynamic Estuary Model, Environmental Protection Agency, Washington, D.C., Water Quality Office, July, 1970.

31. M. B. Fiering and B. Jackson, "Synthetic Streamflows", American Geophysical Union, Water Resources Monograph Series 1, Washington, D.C., 1971.

32. R. A. Freeze, "Three Dimensional, Transient, Saturated Unsaturated Flow in a Groundwater Basin", WRR, Vol. 7, No. 2, 1971.

33. W. E. Gates, R. M. Males and J. F. Walker, "Application of Industrial Dynamics Concepts to Decision Making in Environmental Management", WRR, Vol. 6, No. 6, December, 1970.

34. G. R. Grantham, E. E. Pyatt, J. P. Heaney and B. J. Carter, Jr., "Model for Flow Augmentation Analysis—An Overview", J. of Sanitary Division, ASCE, Vol. 96, No. SA5, p. 1045, October, 1970.

35. G. Graves, et. al., "Mathematical Programming for Regional Water Quality Management", California University, Los Angeles, Graduate School of Business Administration, August, 1970.

36. G. Graves, D. Pinery and A. Whinston, "Application of a Large Scale Non-Linear Programming Algorithm to Pollution Control", Krannert Graduate School of Industrial Administration, Lafayette, Ind., February, 1971.

37. W. Grayman, et. al., "Integrated Development Plan for the Rio Colorado, Argentina, 51st Annual Meeting, American Geophysical Union, Washington, D.C., October, 1971.

38. W. Grayman and P. S. Eagleson, "Evaluation of Radar and Raingage Systems For Flood Forecasting", R. M. Parsons Lab. for Water Resources and Hydrodynamics, M.I.T. Report No. 138, August, 1971.

39. Y. Y. Haimes, "Modeling and Control of the Pollution of Water Resources Systems via Multilevel Approach", Water Resources Bulletin, AWRA, Vol. 7, No. 1, pp. 93–101, February, 1971.

40. I. Haissman, "Generating Skilled Manpower for Irrigation Projects in Developing Countries: A Study of Northwest Mexico", WRR, Vol. 7, No. 1, February, 1971.

41. D. Haith, "The Political Evaluation of Alternative Metropolitan Water Resource Plans", Water Resources Center, Cornell University, Ithaca, New York, July, 1971.

42. H. R. Hamilton, D. E. Boyce, F. J. Cesario, S. E. Goldstone, D. C. Sweet, et. al., "A Dynamic Model of the Economy of the Susquehanna River Basin, Final Report", Battelle Memorial Institute, Columbus, Ohio, August, 1966.

43. S. H. Hanke, "Dynamic Aspects of Urban Water Demand", Proc. of the 4th American Water Resources Conference, AWRA, New York, New York, pp. 642–650, November, 1968.

44. B. M. Harley, F. E. Perkins and P. S. Eagleson, "A Modular Distributed Model of Catchment Dynamics", Report 133, Ralph M. Parsons Lab., Department of Civil Engineering, M.I.T.

45. J. E. Hass, "Optimal Taxing for the Abatement of Water Pollution",

Water Resources Research, Vol. 6, No. 2, pp. 353–365, April, 1970.

46. M. Herdari, V. T. Chow, D. D. Meridith, "Water Resource Systems Analysis by Discrete Differential Dynamic Programming", Hydraulic Engineering Series No. 24, University of Illinois, January, 1971.

47. C. S. Hollings, "Ecologic Models: A Status Report," Proc. International Symposium on Modeling Techniques in Water Resource Systems, Volume 1, Environment Canada, Ottawa, Canada, May, 1972.

48. P. House and P. D. Patterson, Environmental Laboratory for the Social Sciences, U.S. Environmental Protection Agency, GPO, 1972.

49. C. W. Howe, and F. P. Linaweaver, Jr., "The Impact of Price on Residential Water Demand and Its Relation to System Design and Price Structure", Water Resources Research, Vol. 3, No. 1, pp. 13–22, February, 1972.

50. J. Y. C. Huang and E. R. Baumann, "Least Cost Sand Filter Design for Iron Removal", Journal of the Sanitary Engineer, ASCE, Vol. 97, SA2, April, 1971.

51. Hydrologic Engineering Center, "1971 Annual Report", U.S. Army Corps of Engineers, 609 Second Street, Davis, California.

52. H. Jacoby and D. P. Loucks, "Combined Use of Optimization and Simulation Models in River Basin Planning", WRR, Vol. 8, No. 6, December, 1972.

53. D. L. Jameson, "A Model Relating Water Quality, Vegetational Structure and Urbanization in the San Jacinto River Basin", Water Resources Research.

54. E. Joeres, J. Liebman, and C. S. ReVelle, "Operating Rules for Joint Operation of Raw Water Sources", WRR, Vol. 7, No. 2, April, 1971.

55. E. Kalley, "Automatic Planning of the Least Cost Water Distribution Network", Water and Water Engineering, Vol. 75, No. 902, April, 1971.

56. D. Kleinecke, "Use of LP for Estimating Geohydrologic Parameters of Groundwater Basins", Water Resources Research, Vol. 7, No. 2, April, 1971.

57. P. Kirshen, D. H. Marks and J. C. Schaake, Jr., "Mathematical Model for Screening Storm Water Control Alternatives", M.I.T. Ralph M. Parsons Laboratory for Water Resources and Hydrodynamics, Report No. 157, October, 1972.

58. A. V. Kneese and B. T. Bower, Eds., Managing Water Quality: Economics, Technology, Institutions, Johns Hopkins Press, Baltimore, Maryland, 1968.

59. C. A. Kohlhass and D. E. Math, "An Algorithm for Obtaining Optimal Looped Pipe Distribution Networks", 6th Annual Symposium on the Application of Computers to the Problems of Urban Society, ACM, October, 1971.

60. J. A. Lager, E. E. Pyatt, and R. R. Shubinski, "Storm Water Management Model", Vol. I-IV, Water Pollution Control Research Series 11024, DOC 07/71, EPA, 1971.

61. J. Liebman and W. Lynn, "The Optimal Allocation of Stream Dissolved Oxygen", Water Resources Research, V2, No. 3, 1966.

62. J. C. Liebman, "A Heuristic Aid for the Design of Sewer Networks, J. of Sanitary Engineering, ASCE, Vol. 93, No. SA4, 1967.

63. J. C. Liebman and D. Marks, "A 'Balas' Algorithm for Zoned Uniform Treatment", Proc. of the National Symp. on Estuarine Pollution, Stanford University, Stanford, California, August, ASCE, Oct., 1968.

64. E. M. Lofting and H. C. Davis, "The Interindustry Content Matrix: Applications on a Multiregional Basis", WRR, Vol. 4, No. 4, August, 1968.

65. E. M. Lofting, "Economic Evaluation of Water, Part VI: A Dynamic Interregional Input-Output Programming Model of the California and Western States Economy", Water Resources Center, Contribution No. 128, University of California, Berkeley, California, June, 1969.

66. D. P. Loucks, "Computer Models for Reservoir Regulation", J. of Sanitary Division, ASCE, Vol. 94, No. SA4, pp. 657–670, August, 1968.

67. D. P. Loucks, C. S. ReVelle and W. R. Lynn, "Linear Programming Models for Water Pollution Control", Management Science, Vol. 14, No. 4, Series B, pp. 166–181, December, 1967.

68. A. Maass, et al., "The Design of Water Resource Systems", Harvard University Press, 1962.

69. D. C. Major, "The Impact of the Systems Approach on the NAR Study", 52nd Meeting, AGU, April, 1971, WRR, 1972.

70. D. Major, "Benefit-Cost Ratios for Projects in Multiple Objective Investment Programs", WRR, Vol. 5, No. 6, December, 1969.

71. B. B. Mandelbrot, and J. R. Wallis, "Noah, Joseph and Operational Hydrology, WRR, October, 1968, Vol. 4, No. 5.

72. B. B. Mandelbrot and J. R. Wallis, "Some Long-Run Properties of Geophysical Records", Water Resources Research, Vol. 5, No. 2, pp. 321–340, April, 1969.

73. S. Marglin, "Public Investment Criteria", M.I.T. Press, Cambridge, Mass., 1967.

74. N. C. Matalas, "Optimum Gaging Station Location", Proc. of the IBM Scientific Computing Symposium on Water and Air Resource Management, New York, October, 1967.

75. W. Miernyk, "An Inter Industry Forecasting Model with Water Quantity and Quality Constraints—Systems

Analysis for Great Lakes Water Resources", Proc. 4th Symp. on Water Resources Research, Water Resources Center, Ohio State University, Columbus, Ohio, 1969.

76. T. L. Morin, and A. Esogbue, "Some Efficient Dynamic Programming Algorithms for the Optimal Sequencing and Scheduling of Water Supply Projects, WRR, Vol. 7, No. 3, June, 1971.

77. O'Neill, et. al., "A Preliminary Bibliography of Mathematical Modeling in Ecology", ORNL–IBP–70–3, Oak Ridge National Laboratory, October, 1970.

78. L. Ortolano and H. A. Thomas, Jr., "An Examination of Non-Treatment Alternatives in Water Pollution Control", Harvard Water Program, Harvard University, Cambridge, 1968.

79. B. C. Patten, Ed., "Systems Analysis and Simulation Ecology", Academic Press, N.Y., 1971.

80. G. D. Pence, J. M. Jeglic and R. V. Thomann, "Time-Varying Dissolved Oxygen Model", J. of the Sanitary Engineering Div., Vol. 94, No. SA2, ASCE, Proc. Paper 5915, pp. 381–402, April, 1968.

81. G. F. Pinder and J. D. Bredehoeft, "Application of Digital Computers for Aquifer Evaluation", WRR, Vol. 4, No. 5, pp. 1069–1094, October, 1968.

82. D. L. Reddel and D. K. Sunda, "Computer Simulation of Waste Transport in Groundwater Aquifers: Groundwater" Colorado State University, Fort Collins, Natural Resources Center, June, 1969.

83. C. S. ReVelle, D. P. Loucks and W. Lynn, "Linear Programming Applied to Water Quality Management", W.R.R., Vol. 4, February, 1968.

84. ———, "A Management Model for Water Quality Control", J. of Water Pollution Control Federation, Vol. 39, No. 7, pp. 1164–1183, July, 1967.

85. C. S. ReVelle, E. Joeres and W. Kirby, "The Linear Decision Rule in Reservoir Management and Design: 1. Development of the Stochastic Model", WRR, Vol. 5, No. 4, pp. 767–777, August, 1969.

86. C. S. ReVelle, and W. Kirby, "The Linear Decision Rule in Reservoir Management and Design: 2. Performance Optimization", WRR, Vol. 6, No. 4, pp. 1033–1044, August, 1970.

87. T. G. Roefs, "Reservoir Management: The State of the Art", IBM Washington Scientific Center, Report No. 320–350, Wheaton, Maryland, July, 1968.

88. P. P. Rogers and D. V. Smith, "The Integrated Use of Ground and Surface Water in Irrigation Project", American J. of Agricultural Economics, Vol. 52, No. 1, pp. 13–24, February, 1970.

89. P. P. Rogers, "A Game Theory Approach to the Problems of International River Basins", Water Resources Research, Vol. 5, No. 4, pp. 749–760, August, 1969.

90. R. F. Scarato, "Time-Capacity Expansion of Urban Water Systems", Water Resources Research, Vol. 5, No. 5, pp. 929–936, October, 1969.

91. J. C. Schaake, Jr., "A Model for Estimating Regional Water Needs", 52nd AGU Meeting, April, 1971, Washington, WRR, 1972.

92. P. Shiers and D. H. Marks, "A Thermal Pollution Abatement Evaluation Model for Power Plant Siting", Report to the National Science Foundation, M.I.T., Energy Laboratory, Cambridge, February, 1973.

93. R. Smith, "Preliminary Design and Simulation of Conventional Wastewater Renovation Systems Using the Digital Computer", Federal Water Pollution Control Administration, Report No. WP–20–9, Cincinnati, Ohio, 1968.

94. M. J. Sobel, "Water Quality Improvement Programming Problems", Water Resources Research, Vol. 1, No. 4, pp. 477–487, 4th Quarter, 1965.

95. K. D. Stolzenbach, E. E. Adams, and D. R. F. Harleman, "A User's Manual for Three Dimensional Heated Discharge Computations", R. M. Parsons Laboratory for Water Resources and Hydrodynamics, Department of Civil Engineering, M.I.T., TR 156, September, 1972.

96. E. Swanson, "Economic Analysis of Water Use in Illinois Agriculture", Illinois Water Resource Center, Report No. 38, National Technical Information Service, PB-197732, January, 1971.

97. M. G. Tadros and R. J. Kalter, "A Spatial Allocation Model for Projected Outdoor Recreation Demand", WRR, Vol. 7, No. 4, August, 1971.

98. V. J. Tarassov, H. J. Perlis and B. Davidson, "Optimization of a Class of River Aeration Problems by the Use of Multivariable Distributed Parameter Control Theory", WRR, Vol. 5, No. 3, June, 1969.

99. G. S. Taylor and J. Luthin, "Computer Methods for Transient Analysis of Water Table Aquifers", WRR, Vol. 5, No. 1, February, 1969.

100. Texas Water Development Board, "Simulation of Water Quality in Streams and Canals: Theory and Description of the QUAL–1 Mathematical Modeling System", Report No. 118, EPA–OWP–TEX–128, May, 1971.

101. Texas Water Development Board, "DOSAG—1 Simulation of Water Quality in Streams and Canals, Program Documentation and Users Manual", EPA–OWP–TEX–DOSAG–1, September, 1971.

102. Texas Water Development Board, "QUAL–1 Simulation of Water Quality in Streams and Canals, Program

Documentation and Users Manual", EPA-OWP-TEX-QUAL-1, September, 1970.

103. E. L. Thatcher and D. R. F. Harleman, "A Mathematical Model for the Prediction of Unsteady Salinity Intrusion in Estuaries", R. M. Parsons Lab. for Water Resources and Hydrodynamics, Department of Civil Engineering, M.I.T., TR 144, February, 1972.

104. Tracor, Inc., "Estuarine Modeling: An Assessment", Water Pollution Control Research Series 16070 DZV, EPA, UGPO, February, 1971.

105. G. Tschannerl, "Designing Reservoirs with Short Streamflow Records", WRR, Vol. 7, No. 4, August, 1971.

106. C. Upton, "Application of User Charges to Water Quality Management", WRR, Vol. 7, No. 2, April, 1971.

107. C. Upton, "A Model of Water Quality Management under Uncertainty", WRR, Vol. 6, No. 3, June, 1970.

108. A. Wolman, "The Metabolism of Cities", Scientific American, September, 1965.

109. G. K. Young, "Finding Reservoir Operating Rules", J. of Hydraulics Division ASCE, Vol. 93, No. HY6, November, 1967.

110. G. K. Young, J. C. Moseley and D. E. Evenson, "Time Sequencing of Element Construction in a Multi-Reservoir System", Water Resources Bulletin, AWRA, Vol. 6, No. 4, July-August, 1970.

111. G. K. Young and L. T. Gitto, "Stream Flow Regulation for Acid Control", IBM Symp. on Water and Air Resources Management, October, 1967.

Chapter 5

Models in Solid Waste Management

By

Jon C. Liebman

SUMMARY 141

I. INTRODUCTION 141

- Models of Policy Decisions 143
- Models of Management Decisions 143

II. USE OF MODELS IN SOLID WASTE MANAGEMENT 144

- Models Relating to Waste Materials 144
- Models Relating to Fixed Facilities 146
- Models Relating to Vehicles 149
- Models Relating to Manpower 152
- Models Relating to Entire Systems 153
- Miscellaneous Models 154

III. EXAMPLES OF MODELS IN SOLID WASTE MANAGEMENT 155

- A Transfer Facility and Site Selection Model 155
- A Collection System Simulation Model 157
- The Jacksonville Study 160

IV. USE OF OPTIMIZING AND SIMULATION MODELS 161

REFERENCES 162

Preceding page blank

Models in Solid Waste Management

SUMMARY

The management of the urban solid waste system is among the most complex municipal or regional governmental tasks, principally because of the wide diversity of the components of solid wastes and the variety of systems in existence. Modeling is therefore frequently done on an ad hoc basis, and the number of general models which can be used in different situations is relatively small.

In this chapter, we first discuss the general nature of solid waste management. Decisions are categorized as long-range policy decisions and management decisions; the focus of the chapter is on the latter, primarily due to the paucity of successful models dealing with large scale resource policy in the solid waste area. We identify the two-stage nature of management decisions, involving the specification of desired service and the determination of solutions which provide the service specified.

Models which deal with the components of the solid waste system are explored in detail. These include predictive models for waste generation, optimization models for planning the installation of fixed facilities, models dealing with routing and scheduling of the collection vehicles, and models which consider the scheduling of manpower. A subsequent section considers simulative models of multiple components or entire systems.

The final section of this chapter discusses the detailed use of two solid waste management models: an optimization model of transfer station site selection, and a simulation model of an entire solid waste collection system. These two models have been applied jointly in the city of Baltimore, and one of them has also been used in a study in Jacksonville, Fla. We describe their use in these studies, and the kind of information which they were able to provide for decision-making.

I. INTRODUCTION

Solid waste management problems are among the most diffuse of governmental problems. Until quite recently, no need has been perceived nor effort made to view the solid waste system as a single entity. The problems which are now recognized cross political boundaries at all levels and, in many cases, reach outside the traditional spheres of governmental activity at any level.

There are a number of properties of solid wastes which compound the difficulties of decision-making and modeling. The first is that the term solid waste encompasses an almost incredible variety of materials. Ordinary household wastes, which in themselves contain many different components, make up only a small part of the problem. Bulky domestic wastes, such as refrigerators, bedsprings, etc., require a different mode of handling. Still within the area of municipal concern are such diverse items as dead animals, demolition rubble, used Christmas trees, and abandoned automobiles. Industrial waste materials in various forms, some relatively innocuous, others hazardous, present yet a different set of difficulties. In essence, solid waste is made up of all materials which are unwanted and which cannot be readily disposed of into the gaseous or liquid waste systems.

Preceding page blank

SOLID WASTE MANAGEMENT

Models Discussed in Chapter

Summary Table

Model/Decision Area	General Type	Important Characteristics
Waste:		
Generation Prediction	Forecasting	Uses historical data to forecast amount of waste generated
Generation Prediction	Input-output	Permits examination of impact of changes in one sector on waste stream in other sectors
Fixed Facilities:		
Site Selection	Integer programming	Selects sites for facilities from among specified alternatives; considers costs of transportation, construction, and operation
Capacity Expansion	Optimizing: integer or dynamic programming, heuristic	Usually neglects changes in land value, interest rates
Facility Operation	Queueing	Analysis number of loading docks, size of storage facilities.
Vehicles:		
Vehicle Replacement	Integer linear programming	Selects vehicles to be replaced
Single Vehicle Routing	Travelling salesman, truck dispatching	Requires specific collection points
Single Vehicle Routing	Chinese postman	Treats collection as continuous along streets
Multiple Vehicle Routing	Heuristic	Simultaneously districts a large area into truckloads and routes individual vehicles
Multiple Vehicle Routing	Simulation (random walk)	Determines routes randomly; user selects best route found.
Manpower:		
Crew Assignment	Non-linear programming heuristic	Sequences vacation, time off, and overtime
Route Scheduling	Linear programming	Minimizes overtime and penalties for late or early services
Overall System:		
System Operation	Simulation	Permits exploration of effects of policy and equipment change

Most solid waste problems have been dealt with in local areas near the source of the wastes. This is, in part, true because of the very nature of solids—they don't go anywhere. Liquid and gaseous wastes move by the force of gravity, winds, diffusion; they have an effect on people at some distance from their source and, as a result, bring about more centralized, regional activity aimed at the mitigation or elimination of the damage they cause. Solids simply lie where they are discarded. Historically, concerted action has been taken when solid wastes have been viewed as causing community or regional problems; thus, garbage collection was undertaken as a governmental activity out of concern for public health and aesthetic deterioration of neighborhoods, while the unseen accumulation of waste material from an industry has often been left as a problem to be handled by the individual firm. Society has regularly taken the position that unwanted materials should be removed only as far as necessary to get them out of the way. This accounts for the burgeoning of automobile graveyards, which was permitted until they were perceived as a blight upon the landscape, interfering with the public welfare.

This local nature of the problem has naturally led to a wide variety of local solutions. Each individual producer of waste, each firm, each municipality has developed its own method of handling its own particular problems. There has been little motivation, either economic or political, to work together for common solutions. As a result, there are wide differences in the methods of dealing with solid wastes, even between similar municipalities. There are two important effects of these local differences. First, large existing capital investments in fixed facilities in each locality make the development of regional solutions, which would be advantageous if starting from scratch, frequently uneconomical. Second, models and techniques which are useful in improving the situation in one locality may be quite useless in another place because of the differences within the existing systems.

A related difficulty which hampers modeling efforts is the sparseness of usable data. It has become quite clear in recent years that large amounts of data are available. But each unit involved in some aspect of solid waste handling gathers its data somewhat differently. Bookkeeping techniques are different. Collection vehicles, even though they may be individually weighed, collect different components of solid waste in different cities, so that it is difficult to develop a true picture of the makeup of waste streams. Labor practices, particularly methods of payment, vary so greatly as to make it virtually impossible to generalize on labor costs from one place to another.

It should be clear from the above paragraphs that there is no unified view of the solid waste system. There is no single plan of attack; there are no prescriptions for a planning and decision-making process. Thus, to discuss decision-making in solid waste management automatically means a fractionated approach which looks at various aspects of the overall problem, always seeking, of course, to integrate into larger and larger assemblages.

Models of Policy Decisions. It is convenient at the outset to recognize two distinct kinds of decisions. One, which will be discussed only briefly in this chapter, might be termed the long-range, large-scale policy decision. Included in this class are decisions relating to national resource management policy; e.g., the extent to which recycling and reclamation should play a part in solutions of solid waste problems, and the methods to be used to encourage and implement such policies. Modeling efforts in this area have rarely been attempted. The most promising approach is in the use of modified input-output models.

The usual economic input-output model provides information concerning the impact of a change in the output of one sector of the economy on all other sectors. Such a table may be modified to include waste flows as well as money flows. In this form, it provides an estimate of the impact of changes in waste policy on the economy. The results, however, must be used with care, since an implicit assumption of the table is that industries continue to use raw materials in the same proportions (i.e., constant technology).

Models of Management Decisions. The decisions with which the remainder of this chapter deals might best be termed "management" decisions. They are more commonly encountered on local, regional, and occasionally state levels and they are most frequently concerned with the planning and operation of facilities and equipment within an existing institutional and policy framework. This implies that, for the most part, we take as given the generation of solid waste in its current or projected varieties and quantities, and consider only the system used for collecting, transporting, storing, treating, and disposing of this waste. Separation for recycling, and other methods of reclaiming useful materials, may be included within this system and considered on an operational, rather than policy, level.

Within this context, it is helpful to view the decision-making process as a two-stage one, though in fact the two stages usually are carried out jointly. One stage is to specify the kind of service desired, while the other involves determining methods for obtaining that service at least cost. Clearly the cost influences the decisions regarding service quality. In fact, if quantitative measures of benefit could be associated with the parameters of service quality, the separation of decisions would not be necessary and it would be possible to optimize the entire system for maximum net benefits. Unfortunately, this is not the case and the decision maker usually finds himself in the position of exploring the costs of obtaining various levels of service, and finally selecting the desired level subjectively.

Service quality decisions, whose values cannot readily be quantified, include frequency of pick-up, location of pick-up (whether backyard or curbside), kinds of waste which will be picked up, and whether or not separation by the user will be required. These, and similar qualitative questions, are not normally explored by optimization models but various descriptive and predictive models can be used to estimate the costs involved with changing service quality.

Cost minimization questions include number and type of facilities (landfills, incinerators, transfer stations), and their location. Size and type of collection vehicles, and routing of these vehicles, as well as number of men in the crew, also must be decided. This list can be extended, of course, to most of the detailed decisions involved in the complete design of any collection and disposal system.

A solid waste system consists of four components: (1) the waste itself, (2) the fixed facilities used for transfer, treatment and disposal, (3) the vehicles used for transporting the waste, and (4) the men who operate the system. In the following sections, models of each of these components are discussed. Most of these models are optimizing models with an objective of minimum cost. The principal exception is in models of waste generation, which are merely predictive. In a sense, optimizing models of waste production depend on the development of control variables not yet in existence, and are related to the policy decision models discussed in the preceding section.

Following the subsections on models of system components is a discussion of models of entire systems. Here the emphasis shifts from optimization to description and prediction, simply because entire solid waste systems are generally too complex for optimization techniques.

II. USE OF MODELS IN SOLID WASTE MANAGEMENT

Models Relating to Waste Materials. Of particular importance to any analysis of a solid waste system is the ability to predict amounts of waste which will be generated. Depending upon the particular problem being analyzed, waste generation models may be very highly aggregated, predicting amounts for large areas (as, for example, in a study of future disposal facility needs), or they may be very detailed, providing information on a block-by-block basis (as would be needed in a detailed vehicle routing study). Similarly, the time-horizon over which prediction is needed also depends upon the purpose of the study. Waste generation models are rarely needed independent of some other model which uses waste generation as an input.

The simplest solid waste generation model is of the form (Rao, et al. [43]):

$$S_{ij} = \sum_{k} R_k a_{ijk}$$

where S_{ij} is the amount of waste generated in the ith time period in the jth section of the planning region, a_{ijk} is the number of waste generating entities of type k in the ith time period in the jth section of the planning region, and R_k is the unit generation of waste by the kth type of entity. Generating entities may be defined quite crudely (i.e. residential, commercial, and industrial), or they may be much more finely divided. This model assumes that the waste generation rate of any type of generating entity remains constant over time, the generation multipliers R_k being determined by observation of an existing system. A somewhat more complex model permits assumptions about changing rates of generation, and takes the general form:

$$S_{ij} = \sum_{k} R_{ik} a_{ijk}$$

where R_{ik} is the unit generation rate of the kth entity type in the ith time period. Obviously, obtaining data for such a model is much more difficult.

The shortcomings of models such as the above are obvious. They assume that historical data contain all the relevant information, and as a result, permit little analysis of proposed changes in waste generation systems.

Some efforts have been made to model waste generation, including more detailed information about types of wastes generated, by the use of regional input-output models. The data requirements for such models are enormous and data collection constitutes a major project in itself. The advantage of input-output models is that they permit examination of the impact of a change in operation of one economic sector of the model on solid waste generation throughout the entire system. As suggested earlier, their shortcoming is that they do not consider technology changes inside an industry as the result of changes in other portions of the system. For example, the introduction of a major paper recycling industry which purchases waste paper might cause other industries (or households) to shift in the direction of more waste paper production; the input-output model assumes a constant rate of waste paper production per unit of output of the industry.

Two other important shortcomings of all waste generation models should be mentioned. The first is that there is some degree of dependence in every such model on projections of population over the time horizon; recent history demonstrates clearly that the art of population projection, despite significant advances, has a long way to go. When one adds to this the large effects of technology change on amounts and types of solid waste, one must conclude that any projection for other than short time periods must make large allowances for error.

A second, somewhat unsettling factor is that we do not completely understand solid waste generation, even in the presumably simple household waste system. This is nowhere more clearly demonstrated than in the statistical study by Quon, Tanaka, and Charnes [41]. In this project, household generation rates were studied before and after a shift from once-per-week to twice-per-week collection. Control areas, in which a shift was not made, were also used. The study demonstrated that there was a very significant increase in the rate of generation after the change and that this increased rate persisted, even a year later. The implication for a system manager considering the costs of a change in collection frequency is somewhat frightening. One is also tempted to wonder what other apparently remotely related factors have large effects on generation rates.

This is not intended to imply that predictions of waste generation rates are quite as unreliable as predictions of heights of women's hemlines; but neither are they as cut-and-dried as predictions of the time of tomorrow's sunrise. Waste generation projections are required for many management decisions and modeling is as rational an approach as we currently have available. There is a clear need for substantial additional study in this area.

Papers in the bibliography which deal with waste generation models are Golueke [18], Golueke and McGauhey [19 and 20], Morse and Roth [37], and Quon, Tanaka, and Charnes [41].

Models Relating to Fixed Facilities. There is a number of problems related to fixed facilities which are amenable to investigation by means of models. Two of these which frequently are solved by very similar models are the selection of types of facilities to use for treatment (facility selection problems), and the selection of locations at which to install these facilities (site selection problems). The ability to consider these two problems with similar models, and even simultaneously within the same model, stems from the fact that two different types (or sizes) of facility which might be installed at the same site can simply be considered as two different sites, with different associated costs, and thus treated within a general site selection model. In addition to facility and site selection, models may also be used to explore problems of capacity expansion over time and to investigate possible operating rules for the existing facilities.

Models of site selection problems have a long history, dating back to work by Alfred Weber [59]. The fundamental premise of all such models is that facility location depends on finding the optimal balance between costs of building and operating facilities and the costs of transporting material to and/or from these facilities. Some of the early work in site selection assumes that only one (or, in some cases, a fixed number greater than one) facility is to be built and that the cost of constructing and operating the facility is the same wherever it is placed. In this case the site selection problem is simply to determine the location(s) at which to build, such that transport costs are minimum.

A particularly simple solution for the site selection problem is available in the (rather unrealistic) case of linear costs of facility capacity and linear transportation costs. That is, the model assumes that it costs D_j dollars for each ton of waste which passes through an incinerator at site j, and T_{ij} dollars for each ton of waste to be shipped from source (collection area) i to site j. Given a set of potential sites for facilities, and a set of sources (collection areas) of known amounts of waste, the model is:

Minimize:

$$\sum_i \sum_j (T_{ij} + D_j) X_{ij}$$

Subject to:

$$\sum_j X_{ij} = a_i \quad \text{for each source } i$$

where X_{ij} is the amount of waste shipped from collection area i to site j, and a_i is the amount of waste generated at collection area i. The objective function is simply the total cost, including both transportation and facility costs, and the constraints require that the total amount of waste generated in each collection area be collected. This problem is a simple linear program. One may also add upper limits on the capacity of the facility to be constructed at site j, in the form

$$\sum_i X_{ij} \leq b_j \quad \text{for each site } j$$

The problem may be solved by use of the standard Hitchcock transportation method, which is even more efficient than linear programming. The solution provides the amount of waste shipped to each potential site (and, thus, the capacity required for each facility), and allocates the waste from each source to the appropriate facilities. Facilities which receive no waste are, of course, not built. A similar approach may also be used to include intermediate facilities such as transfer stations where waste is transferred from small collection vehicles to larger long-haul vehicles for transport to a distant treatment or disposal facility; the resulting model is in the form of a transshipment problem which may also be solved by a well-known technique.

The principal difficulty with these

models is that they neglect the initial cost of establishing a facility. Because of this usually rather large cost, the unit costs for a facility decrease with its size, and it becomes more attractive to construct fewer, larger facilities. Thus, models of the above type normally overestimate the optimal number of facilities.

There is in existence a number of site and/or facility selection models which overcome this difficulty. By including a 0–1 variable which indicates whether a facility is constructed or not, they are able to consider the fixed costs directly. However, they become mixed integer programming problems and are therefore much more difficult to solve.

The general mathematical form of such models is:

Minimize:

$$\sum_i \sum_j (T_{ij} + D_j)X_{ij} + \sum_j F_j Y_j$$

Subject to:

$$\sum_j X_{ij} = a_i \quad \text{for each source } i$$

$$b_j Y_j - \sum_i X_{ij} \geq 0 \quad \text{for each site } j$$

$$Y_j = \{0,1\}$$

where Y_j is equal to 1 if a facility is built at site j and zero otherwise, F_j is the fixed cost of constructing such a facility, and the other variables are as previously defined. The objective function now includes the fixed cost of a facility which is paid only if $Y_j = 1$ (that is, the facility is constructed). The second constraint requires that if a facility is built, the amount flowing into it may not exceed its capacity, while if it is not built there may be no flow into it.

A very similar model may be used to select among various kinds of facilities or various capacities of the same kind of facility at the same site. For example, suppose that a 500 ton capacity incinerator, an 800 ton capacity incinerator, or a 1500 ton capacity incinerator may be built at a particular site. The three facilities are treated as though they were at three different sites, and the model used is identical to that shown above. However, in order to prevent a solution which constructs more than one of the facilities, a constraint must be added. If the three different facilities are represented by $j = 2$, 6, and 9, the new constraint is

$$Y_2 + Y_6 + Y_9 \leq 1$$

which prohibits building more than one of them. Similar constraints can be used to prevent any particular combination of facilities which, for any reason, is considered impossible or undesirable.

In some cases, these more complex integer programming models are not necessary in order to obtain an optimal solution. If the number of potential sites or facilities is small, it is possible to find the solution by enumeration. Suppose, for example, there are only four possible sites under consideration for the location of incinerators. The optimal solution may have only one facility, or it may have two, or three, or all four. There are four possible solutions with one facility, six possible solutions with two facilities, four possible solutions with three facilities, and one solution with four facilities, for a total 15 possible solutions. For each of these possible configurations, the fixed cost of the facilities is known and it is only necessary to allocate waste sources to the facilities in such a way as to minimize cost. This problem is again in the form of a standard Hitchcock transportation problem and can be solved with comparative ease. Thus, the overall optimum is found by solving 15 transportation problems, one for each of the possible configurations, and selecting the configuration which has the minimum sum of fixed costs plus transportation costs. As the number of potential sites increases, the number of configurations increases exponentially (for n potential sites, the number of configurations is $2^n - 1$), so that the enumerative approach cannot be used for large problems. For example, with five potential sites the number of configurations is 31, with 10 potential sites the number of configurations is 1023, with 20 potential sites the number of configurations is 1,048,575, and with 50 potential sites the number of configurations is approximately 10^{15}.

A particularly important approach to the solution of site selection models is based on a heuristic fixed-charge algorithm by Walker (57). The method is designed to handle any linear programming problem in which there is an initial fixed charge for any variable which becomes non-zero, as well as a linear charge as the variable increases in value. The solution technique is a modification of the simplex method for solving linear programming problems, which does not guarantee global optimality. However, computational experience with the Walker algorithm has demonstrated that it almost always does find the optimum solution, and when it fails it still comes quite close in most cases. The Walker algorithm has been incorporated into a solid waste facility location model by Roy F. Weston, Inc., and further extended by Argonne National Laboratory's Center for Environmental Studies. The code will solve problems with in excess of 100 waste sources and 50 facility sites with ease. It also includes a number of optional features which are particularly appropriate for solid waste problems.

A major criticism of site selection models is that they generally consider only transportation costs and the cost of constructing and operating the facilities. Site selection, particularly for undesirable facilities, actually involves many considerations which can be put into economic terms only with great difficulty if at all. Thus, it is often claimed that site selection models are essentially useless because they can shed no light on these nonquantifiable factors. This attitude is contrary to the philosophy that the purpose of models is to provide useful information to decision-makers. The useful information provided need not be complete answers but must be in a form which can be integrated into the decision-maker's intuitive understanding of the entire problem. In the case of site selection this use of models is particularly easy. Almost every site selection model considers only a discrete set of potential sites which have been identified by the user. Thus, any sites which are felt to be infeasible regardless of the economies they offer can be excluded from the problem at the outset. After the problem solution has been obtained, it is still possible to exclude a questionable site and re-solve the problem, thus obtaining a second-best solution (from the standpoint of cost). The difference in cost between the two solutions provides the decision-maker with a measure of the savings which may be obtained by using the questionable site or, alternately, of the cost incurred by avoiding it. The final weighting of the unquantifiable factors in such problems rightly belongs in the political sphere; but the availability of this cost information provides the political decision-making body with an additional and valuable tool for reaching its decisions.

In one important sense, site selection models represent a very short-sighted view of the overall problem. These models are solved at some particular time when a need for additional facilities is perceived to be sufficiently great as to justify the facilities and they presuppose a decision as to the total amount of additional capacity which is required. In a broader sense, the problem is really one of determining well in advance when new facilities will be needed and how much capacity should be provided at each point in time over some planning horizon. This problem, known as the capacity expansion problem, can also be explored by means of models.

The objective in a capacity expansion problem is to determine optimal size of plants to be constructed (or of additions to existing plants) and the times at which construction is to be undertaken so as to minimize the present value of all future costs. Various assumptions may be made as to the length of the time horizon and time may be treated either as a continuous variable or as a discrete variable. It is, of course, assumed that amount of waste generated is known as a function of time.

Capacity expansion models are generally quite complex and very much specialized to a particular set of assumptions. This specialization, plus the difficulty of obtaining satisfactory data on costs and projected waste generation rates, has hampered widespread use of capacity expansion models in the solid waste field. An additional difficulty is caused by the fact that land costs (and thus, facility costs) change in an urban area as it grows so that the cost of an additional unit of capacity (particularly with respect to landfills) is different at different time periods. Most capacity expansion models do not take this factor into consideration adequately. A particularly extensive model which includes both initial site selection and capacity expansion has been presented by Esmaili [16]. The model uses an elaborate objective function which includes both capital and operating costs of facilities, haul costs, and a discount factor for facilities which are not used for the total period of their useful life. The solution technique proposed appears to be an orderly enumeration of possible configurations for each time period with consideration being given to a particular configuration only if it yields an improvement over the best configuration found so far.

One other application of models to problems involving facilities is worthy of note. Once a facility is in existence, there are likely to be operating policy questions which require investigation. For example, is it more economical to have a very small number of loading docks at an incinerator, thus requiring vehicles to wait for some period of time before unloading, or to have more unloading facilities in order to reduce the waiting time? Similarly, is it more economical to have a large storage area and operate a treatment facility around the clock, or to have a smaller storage facility and operate the treatment facility only during normal working hours? Models for exploring optimal policies of this sort generally fall into the area of queueing theory and one study of the application of such models to solid waste facilities has been made (Stern [51]). As in the case of capacity expansion problems, these models tend to be rather complex and somewhat specialized.

Papers in the bibliography which deal with facility models are: Anderson and Nigam [3]; Clark and Helms [9]; Esmaili [16]; Helms and Clark [22]; Liebman [27] [28]; Marks and Liebman [32] [33] [34]; Nigam [38]; ReVelle, Marks, and Liebman [44]; Rossman [46]; Schultz [48]; Skelly [50]; Stern [51]; Walker [57]; Weber [59]; and Wolfe and Zinn [60].

Models Relating to Vehicles. Modeling may be applied to several different aspects of the waste transport system. The most common applications of models in this area relate to the routing and scheduling of collection (and occasionally long-haul) vehicles. However, models have also been developed which may be used in scheduling collection in different areas, and in selecting replacement vehicles and scheduling maintenance.

A vehicle replacement model (Clark and Helms [10]) assumes the existence of a particular fleet of vehicles and known operating and amortization costs for each vehicle type within the fleet. The model selects the number of each type of vehicle to purchase, given the total number of vehicles to be replaced. The number purchased is not necessarily the same as the number retired, since vehicles of different sizes may be added to the fleet. Although the problem is an integer programming problem, since number of vehicles must be integer, reasonably satisfactory (but not necessarily optimal) solutions may be obtained by using linear programming to obtain a solution and rounding to nearest integers. An extension of the model is possible to permit determining how many vehicles to replace (and which ones), providing operation and maintenance cost data are available for each vehicle (or, at least, for each age-group within a vehicle type).

The overall problem of collection vehicle routing has received a great amount of attention, perhaps because it is closely related to so many other routing problems (school buses, milk delivery, parcel delivery, mail delivery, etc.). The entire problem requires dividing the collection region into individual areas, each of which generates one truck-load of waste, and then routing a vehicle from a depot to its area, through the area, and back to the depot. Ideally, the division into truck-load areas (called districting) and the routing of individual trucks should be accomplished simultaneously since the two are interdependent. In many modeling efforts, however, the two steps are examined by independent models.

The objective in routing and districting is to provide the required level of service (collection from each waste generator at specified intervals, usually on specified days) with minimum cost. Cost generally consists of two separate items: depreciation of the vehicle, which is a function of distance traveled, speed, and load; and labor cost, which is a function of time. Most models ignore the effect of speed and load on vehicle depreciation, thus making cost a function of distance traveled and time.

Early attempts at modeling collection routing took the form of the classic "traveling salesman" problem. In its original form, the traveling salesman problem requires a minimum distance solution which starts at a specified point, visits a number of points which require service (in any order), and returns to the original point. To obtain an exact optimal solution to the collection vehicle routing problem in this form would require treating each individual waste source (household) as a point which must be visited and even then the solution might propose turning around in the middle of a block, which is generally unacceptable. More frequently, traveling salesman models are used by aggregating individual sources into larger units (say, a block face, or a street segment) which are then treated as points requiring service. Solution methods for determining optimal routes for traveling salesman problems are generally unwieldy for problems with more than about 50 points. A number of rather good heuristic approaches do exist, however. There have been some attempts to extend the traveling salesman approach to obtain routes for more than one "salesman" when all must start and finish at the same depot, but each point need be visited by only one salesman. This is the equivalent of incorporating the districting and routing problems into one model. This approach was pioneered by Dantzig and Ramser [13], and the problem formulation has become known as the Truck Dispatching problem. A more advanced algorithm based on the Dantzig and Ramser method was proposed by Clarke and Wright [11], and is used in modified form in the IBM Vehicle Scheduling Program which is available for System/360 computers. Although the algorithm is intended for cases with a comparatively small number of points requiring service, it has been used successfully to determine good (but not necessarily optimal) routings for large municipal areas.

In recent years, some progress has been made through modeling the collection routing problem in the form of the so-called "Chinese Postman's problem." The assumption in this model is that service is required along all of the streets in a network, rather than at specific points. Thus, the single truck routing problem attempts to find a route through the street network which traverses every street at least once, and minimizes the total distance traveled. It should be clear that a lower bound on total distance is the sum of the lengths of all of the streets since that would be the distance traveled if no streets were traversed more than once. Leonhard Euler, in 1736, showed that it is possible to traverse each street exactly once if and only if the number of streets incident upon each intersection is even. The intuitive basis for this proof is that for every entry into a particular intersection there must be an untraversed exit; thus, each time the vehicle enters and leaves the intersection it utilizes exactly two streets. If the number of streets touching the intersection is odd, there will be one remaining untraversed street after the last exit from the intersection and re-entering the intersection using this street leaves no untraversed street on which to depart. From this it is easy to see that if one wishes to make a complete tour through a network traversing every street at least once, it will be necessary to retrace at least one street from each intersection which has odd degree (degree refers to the number of streets incident upon the intersection). But if the retraced street leads to an intersection which was of even degree, the retracing effectively makes that intersection odd. The only way to prevent a retracing from making an even intersection odd is for the retracing to go between two intersections which are odd. It may be that several streets must be retraced to get from one odd intersection to another; this retraced path both enters and exists from intervening even intersections, thus leaving them even, as shown in Figure 1. Any tour which traverses all streets in a network at least once will retrace paths between pairs of odd intersections; since the objective is the minimization of total distance, it is necessary to "pair up" the odd intersections so that the sum of these paths connecting pairs be minimum. A particularly elegant solution method developed by Edmonds [14] may be used to find the optimum pairs.

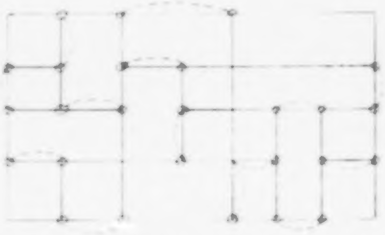


FIGURE 1—*A hypothetical street network, showing retracings necessary in order to construct a tour which traverses all streets at least once. Circles indicate intersections of odd degree; dashed lines show retracings which make the degree of every intersection even.*

Once the streets to be retraced have been selected, there are many different tours through the network which retrace only those streets and the construction of such a tour is a comparatively easy task. Other objectives (such as minimizing left turns) can be considered by the router who constructs the tours.

The problem of routing multiple vehicles (i.e. districting simultaneously with routing) has not been optimally solved. A good but complicated heuristic for the "multiple postman's problem" is given in Edmonds and Johnson [15] but this approach has not been tested in practice.

It should be noted that a simple heuristic solution for the problem can be obtained by treating the entire collection area as if it could be collected by a single vehicle. Once the retraced streets for this fictitious vehicle have been determined, it is possible to draw district boundary lines which pass through intersections in such a way as to keep an even number of streets incident on the intersection on each side of the boundary. Thus, a complete tour is still possible within each district without retracing any more streets than the fictitious single vehicle would retrace. Such an approach, however, neglects the additional retracings which each vehicle will have to make in traveling between its collection district and the depot. A heuristic using this method and considering these added retracings is given by Male [30].

The routing problem may also be modeled by a simulator (Bodner, Cassell and Andros [7]). In such a model, the vehicle begins at an intersection and randomly selects a street on which to collect. Information on the amount of waste collected on that street and the time required may be provided as input data or generated randomly from an assumed statistical distribution; in any case, the amount is added to the truck's load and the time interval is added to elapsed time. When the truck reaches the next intersection it again randomly selects a street on which to proceed. Whenever the truck reaches an intersection where there are no remaining uncollected streets, a search is made for the nearest uncollected street and the vehicle proceeds to this point (retracing) and then begins collecting again. When the truck reaches capacity it returns to the depot. When the elapsed time reaches that of a normal workday, overtime is charged. A route constructed in this manner is not likely to be particularly good; however, the computer time required to run the model is quite short, so it is possible to make a large number of runs and select the best route generated. An improvement in this method might be obtained if the search for nearest uncollected street were converted to a search for the nearest odd intersection with an uncollected street, since paths between odd intersections will ultimately be constructed anyway. The simulator can also be modified to adhere to additional rules which are felt to provide better routes, such as avoiding left turns.

Some criticism has been leveled at attempts to determine optimum vehicle routes (Shuster and Schur [49]). The basis for the criticism is principally three ideas: the data requirements for such models are prohibitive; there are too many properties of a "good" route which are not reflected in a mathematical model; and a man, following a good set of routing rules, can do an excellent job of routing without the use of a complicated model. There is a good deal of validity in these claims, particularly with respect to earlier models which completely determined a route and left no flexibility for the manager. However, the Chinese postman model determines only the streets to be retraced for minimum retracing distance, leaving the manager a great deal of choice in how the tour is actually to be constructed. In effect, as is the case in facility models, the model and the computer are used to do what the manager is poorest at (minimizing distance), while the manager is free to use the excellent rules provided by Shuster and Schur on the resulting network. Though it is true that data requirements are great if exact optimal solutions are desired, such things as travel time for deadheading can be approximated rather readily so that the computer codes can be executed with little more data than the manager already has available. In addition, it is important to note that the trend toward large municipal data bases such as the DIME file of the Bureau of Census will make all the data required for elaborate routing models readily available and make it possible to construct new routes at frequent intervals.

It should be noted that the independence of routing and site selection models is purely artificial. The selection of vehicle routes is clearly dependent upon facility locations; facility locations, however, cannot be truly optimized without regard for the routes which will be generated. Ideally, what is required is a location-routing model which seeks to optimize both simultaneously. In practice, however, routing is usually a comparatively short-term decision, while the location of facilities is considered at infrequent intervals and for much longer time periods. Thus, the use of two independent models probably does not invalidate either.

Papers in the bibliography which discuss vehicle models are: Altman, Bhagat, and Bodin [2]; Andros [4]; Beltrami and Bodin [5]; Bodner, Cassell, and Andros [7]; Brown [8]; Clark and Wright [11]; Coyle and Martin [12]; Dantzig and Ramser [13]; Edmonds [14]; Edmonds and Johnson [15]; Fuertes, Hudson, and Marks [17]; Liebling [26]; Liebman [27] [28]; Lofy [29]; Male [30]; Marks, Cohon, Moore, and Stricker [31]; Marks and Liebman [32]; Marks and Stricker [35]; Martin and Rothwell [36]; Owen [39]; Rothgeb [47]; Shuster and Schur [49]; Stricker [52]; Wathne [58]; and Wyskida and Gupta [61].

Models Relating to Manpower. The most pressing problem in the area of manpower is the scheduling of the workload so that it is equitably distributed among the workers and so that the right number of workers is employed. There are three fundamental steps involved in scheduling. First, it is necessary to determine just what constitutes a day's workload. How many dwelling units (or tons) can a vehicle and its crew collect in an hour? How does crew-size relate to this figure? What is the trade-off between deadheading time and amount to be collected? Modeling yields little information in this area; though some work has been done on the problem, much remains to be determined before models can be used to solve the rest of the questions. Second, the vehicle routes must be balanced so that each crew comes close to the pre-determined fair workload. Heuristic routing models for the multi-truck problem usually include some effort to balance workload but, again, there is more work needed to obtain satisfactory solutions. Third, crews must be assigned to daily routes, and scheduled so that vacation periods, days off, and overtime (if any) are equitably distributed and arranged in satisfactory sequences. This problem has, until recently, received little attention; but several new papers (Altman *et al.* [1], Bodin [6]) have developed a rather complex non-linear programming and heuristic algorithm for scheduling of this sort. Another model (Tanaka and Quon [53]) takes as its objective the short-run scheduling of routes using available manpower so as to minimize the sum of overtime costs and penalty costs associated with late or early service. To date these models have not been sufficiently tested in practice to determine their utility.

Models Relating to Entire Systems. The obvious difficulty with most of the models discussed in earlier sections is that they do not consider the entire solid waste system of a municipality. Thus, the interplay between waste generation, routing of trucks, location and type of facility, and scheduling of manpower is not explicitly considered. Most of the models discussed have been optimizing models; the problem is that any complete solid waste system is too complex to be fitted into the form of an optimization model. Thus, although models of entire systems do exist, they are usually simulation (prediction) models rather than optimization models, permitting the exploration of particular policies and system configurations, but not allowing direct optimization except by trial and error.

Attempts have been made to build simulation models of waste collection and disposal which are general enough to be used in any city. In general, these models have been only marginally successful. Each municipal system has its own quirks and idiosyncrasies, and it is extremely difficult for the modeler to imagine all such individual characteristics and allow for them in his model. Thus, although efforts should be (and will be) continued toward a general model, it is to be anticipated that the best simulation models of overall systems will be tailor-made for the specific system for which their use is intended for some time to come.

System simulators may be constructed at almost any level of detail which is desired, depending on the specific uses for which the models are intended. For example, some models simulate flow of material into and out of facilities; others include simulation of the movement of individual vehicles, while the most detailed might (at least in theory) simulate generation of waste material at individual households and its detailed movement through the system. Generally this most detailed level is not required for the decisions in which the model is intended to assist.

Most system simulators include consideration of the stochastic element of the system, a factor which is not normally included in the optimization models discussed above. In optimal routing models, for example, average vehicle speed is usually used to measure the cost of travel over a particular street. In simulation models, a mathematical or empirical distribution of truck speeds may be provided to the model, and values for speed may then be selected randomly from this distribution whenever they are needed. Thus, multiple runs of a simulator provide different answers just as multiple days of operation of the real system yield different results.

The most common output from a system simulation is a value for total cost of operation of the system. However, depending on the level of detail built into the model, almost any information which could be obtained from observation of the actual system can also be obtained from the simulator. Number of overtime hours worked, number of trips to the incinerator, time of day of collection at specific points, size of queues at transfer stations, and other measures of quality of service can all be obtained.

Again, depending on their level of detail, system models may be used for evaluating and comparing changes in collection frequency, changes in routing assignments and policies, new locations of disposal and treatment facilities, potential locations of transfer facilities, or almost any other change in operation of a system. Such explorations cannot be made in a vacuum, however; the data requirements for simulators are usually quite extensive. Particularly when using a simulation model for evaluating a new component of the system, it is often difficult to obtain necessary information on the performance of the component. It should be kept in mind that simulation models are best used for exploring the interrelationships in a large system of a number of individual components when the behavior of the individual components is well understood. Thus, for example, a study of the need for and effect of transfer stations in a system which does not have such stations cannot be effective unless data are available concerning the performance of the long-haul vehicles which will be used in conjunction with the stations. Frequently, data of this sort must be obtained from other systems which already have the components being considered.

Papers in the bibliography which deal with overall systems models are: Helferich, Hoffner, and Gee [21]; Quon, Charnes, and Wersan [40]; Quon, Tanaka, and Wersan [42]; Rao, Richardson, and Wismer [43]; and Truitt, Liebman, and Krusé [54] [55].

Miscellaneous Models. Two other kinds of models which have application to solid waste management are not readily classified in the framework used above.

The technique of gaming has been used in a number of areas for the purpose of training decision-makers in an environment which simulates the management of an actual system without incurring the high costs associated with management mistakes. An important gaming model is APEX, an air pollution game. The success of these models demonstrates their usefulness and similar models could certainly be beneficial in the solid waste management field. Wahi and Peterson [56] have developed a game which uses a computer in an interactive mode. The player (or players) is faced with a particular situation in which some decision must be made and has at his disposal various computer programs for determining shortest paths between points, optimal allocation of waste sources to existing facilities, etc. The player is introduced to the decision situation gradually, in such a way that his confidence in the use of the computer as an aid is built with experience. Klee [24] has also described a solid waste management game, developed by the Office of Solid Waste Management of the U.S. Environmental Protection Agency.

The management of a solid waste system involves balancing of multiple goals, as suggested early in this chapter. The minimization of cost to provide a required level of service is only one part of the process; the more difficult portion is the determination of the desired quality of service and the determination of whether an increase in some aspects of service quality are worth the associated increased cost and decreased level of other aspects of service quality.

The use of decision models as an aid in the weighting of various quality and cost goals so that alternative systems can be evaluated has not been widely explored; Klee [23] [25] has presented the pioneering work in the application of multiple goal techniques to solid waste management.

III. EXAMPLES OF MODELS IN SOLID WASTE MANAGEMENT

A casual perusal of the references at the end of this chapter will reveal that fully 75 percent of the listed papers bear dates of 1970 or later. This is not due to any attempt on the part of the author to weed out earlier papers which may have been outdated; on the contrary, the bibliography represents a nearly exhaustive list of all the modeling work which has been done with solid waste systems in mind, with the exception of a few reports covering a specific problem in a specific place. It is simply the case that attempts to bring modeling methodologies to bear on solid waste problems are quite recent. For this reason, it is not possible to cite in the following sections many examples of the application of general models to solve real-world problems. In most cases, the models have not yet been applied; in those few cases where models have been applied, either the results have not yet been implemented, or there is insufficient information to judge the final outcome. The selection of the two models discussed in detail below was motivated by the fact that both have been used in somewhat hypothetical studies in the same city (Baltimore) and one has been used in an actual study in Jacksonville. These related uses of the models permit comparison and a discussion of how simulation and optimization may be used in conjunction.

A Transfer Facility and Site Selection Model (Marks and Liebman [32] [33] [34]). Marks has developed an optimizing model which determines appropriate locations for transfer facilities when sources of waste and disposal sites are known. The model minimizes the capital and operating costs of the transfer stations plus the transport costs. Transport costs are taken to begin with the trip from collection area to transfer facility or disposal facility; costs of actual collection are not included.

The collection area is divided into a set of **K** collection tracts (representing, ideally, individual truckloads), with an amount of waste S_k generated at tract k. There is a known set J of disposal points and the capacity of disposal point j is D_j. A set of proposed intermediate (transfer) facility sites has been suggested. Potential transfer facility i has a fixed charge F_i, a capacity Q_i, and a unit processing cost V_i per ton of waste. The mathematical statement of the model is:

$$\text{Minimize: } \sum_i F_i y_i + \sum_k \sum_i b_{ki} x_{ki} + \sum_i \sum_j C_{ij} w_{ij} + \sum_k \sum_j d_{kj} z_{kj} \qquad (1)$$

$$\text{Subject to: } \sum_j z_{kj} + \sum_i x_{ki} = S_k \qquad \text{for each tract } k \qquad (2)$$

$$\sum_j w_{ij} = \sum_k x_{ki} \qquad \text{for each transfer site } i \qquad (3)$$

$$\sum_k x_{ki} - Q_i y_i \leq 0 \qquad \text{for each transfer site } i \qquad (4)$$

$$\sum_k z_{kj} + \sum_i w_{ij} \leq D_j \qquad \text{for each disposal facility } j \qquad (5)$$

$$y_i = \{0,1\} \qquad \text{for each transfer site } i \qquad (6)$$

where:

b_{ki} = cost of transporting a unit (ton) of waste from tract k to transfer facility i, including processing cost at facility i (V_i)

C_{ij} = cost of transporting a unit (ton) of waste from transfer facility i to disposal facility j

d_{kj} = cost of transporting a unit (ton) of waste from tract k to disposal facility j

w_{ij} = amount of waste transported

from transfer facility i to disposal facility j

x_{ki} = amount of waste transported from tract k to transfer facility i

y_i = 1 if the ith transfer facility is constructed, and zero otherwise

z_{kj} = amount of waste transported from tract k to disposal facility j

The objective function (1) requires the minimization of the fixed costs associated with those transfer facilities which are constructed (first term) plus transport costs from collection tracts to transfer facilities and processing costs at transfer facilities (second term) plus transport costs from transfer facilities to disposal facilities (third term) plus transport costs from collection tracts to disposal facilities (last term). Equation (2) requires that the amount of waste taken from each collection tract be equal to the amount generated there; while equation (3) requires that the amount into a transfer facility be equal to the amount out. Equation (4) specifies that if a transfer facility is not built ($y_i = 0$) it can handle no waste, while if it is built ($y_i = 1$) it can handle no more than its capacity. Inequality (5) prohibits shipping to any disposal facility an amount greater than its capacity, while constraint (6) requires that a potential transfer facility can only be built or not-built, and cannot be partially built. (It should be noted that the original model, described in the referenced papers, did not explicitly include flow directly from collection tract to disposal facility, but permitted it by the artificiality of constructing a "dummy" transfer facility with zero cost).

Because the model is a mixed-integer linear programming problem, the solution method proposed is a form of branch-and-bound, using a network flow algorithm for solving the individual branch problems. Cases with 40 collection areas, 7 potential transfer facilities, and 2 disposal sites were solved on an IBM 7094 computer in approximately 45 seconds.

Additional constraints may be placed in the model without significantly affecting the solution technique. The two most important of these are (a) a budget constraint, which limits either the amount of capital costs incurred in construction of facilities or the number of facilities to be constructed, and (b) a mutual exclusivity constraint, which prohibits the construction of more than one (or any other number) of a specified subset of facilities. This latter constraint permits the consideration of several different sizes of facilities at the same point, or the choice of no more than one site in a particular neighborhood.

This model has been applied in a study of the collection system in the northwest quarter of the city of Baltimore. The data used were derived from those collected for the simulation model of Truitt *et al.* [55] discussed below. The area studied is predominantly residential, with collection being made twice per week.

The first run of the model was essentially a verification trial to insure that the data had been abstracted without error and represented the collection area reasonably well. No transfer stations were permitted and the model was simply used to calculate weekly collection costs. The result gave a cost of $16,600 per week while the actual system cost is estimated to be approximately 4 percent higher. This difference was considered to be relatively insignificant, thus indicating that the data were accurate.

The model was then used to explore the potential savings associated with transfer stations. Four sites within the collection area and two outside the area (between the area and the disposal site) were selected as the most likely locations for a transfer station and the construction and operation costs for stations of various capacities were estimated. Several runs were made, each with transfer stations of a different capacity. The results indicated that with an optimally located 600 ton/day capacity station, costs were reduced by approximately $700 per week while with a 900 ton/day capacity station the reduction was approximately $900 per week. Stations larger than 900 tons/day had smaller reductions in total cost, though even a 1500 ton station showed a reduction of about $700 per week. Also of interest was the fact that the optimal site for a station was the same for capacities of 900, 1200, and 1500 tons per day but different for a capacity of 600 tons per day. Additional runs were made for three times per week collection and for increases in waste generation rate ranging from 10 percent to 60 percent; all showed the same stability in selecting the identical site alternative.

It was considered to be of particular importance to determine the possible effects of errors in estimating costs or other data. Thus, a number of runs were made in which facility costs, transport costs, collection rates, etc. were varied; these also demonstrated remarkable stability. Transfer stations remained (marginally) economical until the facility fixed cost almost doubled; and the same site was chosen with almost all combinations of data.

A Collection System Simulation Model (Truitt et al. [54] [55]). The system simulator constructed by Truitt as a doctoral dissertation had as its intent the demonstration of the feasibility of utilizing simulation to compare alternative policies of operation within a single city. It was also intended that the simulator itself should be sufficiently general to permit its use for the above purpose in more than one city; that is, flexibility should be built into the program so that differences between cities could be conveniently encompassed. Policy parameters which are variable within the model for purposes of comparison include: location of a single transfer station (if it exists), location of a single disposal site, collection frequency, collection vehicle size, pay scale and overtime pay policy and vehicle use and amortization charges. Other variables which can be changed indirectly include household density and waste generation rate (varied by changing the data base provided on waste generation), season of the year (varied by changing waste generation rates), crew size (varied by changing the vehicle collection rate data).

Two preliminary simulation models were constructed prior to the development of the third and final model. These two models were used to assist in constructing specific portions of the final model and will not be discussed here.

A single run of the system simulator calculates the result of collection activity by a number of collection trucks in an urban area for one week, Monday through Saturday. The area may be composed of many subareas of different population densities. A transfer station may be present in the area, located at a specified point. In this case, the model also simulates the operation of waiting lines at the transfer station and of the long haul tractors and trailers used for transporting the waste from the transfer station to the disposal site. The disposal site is also located at a specified point.

The output from the simulator lists the cost for the week of the operation of collection vehicles, transfer station, and tractor trailers. These costs include amortization, operation and labor. The total tonnage collected is also provided. In addition, mileage traveled by collection vehicles while collecting and while traveling to and from collection areas is provided, as is total mileage traveled by long-haul vehicles. Amounts of overtime generated and average hours in the workday are printed, as are other miscellaneous items such as the time distribution of the lengths of transfer station queues.

The input data for the simulator may be divided into two sections: geographic data and field performance data. Geographic data for the collection area are generated by dividing the entire area up into small tracts such as census tracts, each with a population of less than about 6000 people. For each such tract, the geographic coordinates of the centroid are provided as well as the total number of households in the tract and the population density in households per acre. In addition, travel distance from the disposal site or transfer station to the tract must be provided, unless the user is satisfied with calculation of the distance based on geographic coordinates. Household waste generation rates for each level of population density must also be determined. Geographic data for the transfer station (if any) and the disposal site simply consist of the coordinates of the locations of each.

Field performance and cost data provided to the simulator include values used for the calculation of amortized cost of the transfer facility, hourly charge for tractor trailer and collection vehicle operation, drivers' and laborers' daily pay rate and overtime rate, capacity of collection vehicle and long-haul vehicle, minimum weights to be accumulated in each of these vehicles before they are permitted to empty, and various miscellaneous values which must be used in calculating vehicle and facility performance.

The simulator calculates speed of the long-haul vehicles, collection speeds of the compactor-collection vehicles, and transit speeds of the compactor-collection vehicles by selecting randomly from frequency histograms provided for each of these values. Thus, the speeds of these vehicles must be observed in actual operation and the required histograms constructed. Several different histograms for collection speeds are required, one for each of several different population densities.

The simulator permits the inclusion of an auxiliary compaction unit at the transfer station to compact the waste additionally as it is loaded into the long-haul vehicles. If such a unit is included, data must be provided concerning its cost and the anticipated density of the output waste.

As the simulator begins a day's operations, it sends the required number of collection vehicles into the field to assigned task areas. It calculates the time and distance traveled by each truck to reach its task area, then calculates the collection time and distance traveled until the truck is full. The portion of the task area which is complete is recorded and the truck is then sent back to the transfer station or disposal site. Time and distance are again calculated and the time required for processing the truck at the transfer station is also calculated, including time waiting in a queue if one exists. The simulator continues in this manner until each assigned task area is complete, including calculation of overtime if it is required. The behavior of long-haul vehicles is determined in similar fashion. At the end of the week's operation, costs and other output data are calculated and printed and the run is terminated.

It is important to note that detailed routing for each truck is not included in this model. Times and distances traveled during collection are based on histograms of average collection rates: times and distances traveled from collection area to disposal site are based on distances from the centroids of collection areas to the location of the disposal facility and on histograms of travel speeds.

The entire simulator, including data, consists of approximately 3500 Fortran cards. The simulator has been run on an IBM 7094 in core memory and has been adapted for use on an IBM 1130. The maximum population which may be simulated on the IBM 7094 is approximately 300,000 people; a single run (one week's operation) of such an area on that machine requires less than one minute of computer time.

The completed model was tested and its uses explored using the City of Baltimore as an example. Although extensive data gathering (involving approximately two man-months and the cooperation of many vehicle operators, foremen, and city officials) was undertaken, it must be kept in mind that the purpose of the study was to demonstrate the utility of the model, not to determine specific answers for Baltimore. Thus, some data (for example, travel speeds of long-haul vehicles, which are not currently used in Baltimore) were extrapolated from other cities rather than making short test-runs in Baltimore.

The first runs of the model were made as proof runs to determine how closely the model duplicated existing conditions in the northwest quarter of the city. Two such runs were made; each resulted in total costs within 2 percent of the actual system costs, with all other output falling within a similar distance from observed values. This close agreement is not as remarkable as it might seem since, after all, the underlying data for the model are simply statistical measures of the performance of the actual system. It does demonstrate, however, that the simplifications inherent in aggregating household units into larger tracts and measuring collection speeds for only a few different household densities do not significantly affect the validity of the model.

The model was next used to investigate the effects on cost and other system parameters of increasing collection frequency from twice per week to three times per week. To make runs simulating this collection frequency, it was necessary to adjust the frequency histograms of collection rates to reflect the reduced amount of waste at each collection. Since Baltimore has no areas in which collection is made three times per week, no data were available for the construction of these adjusted histograms. However, in the existing, twice-per-week collection system, collections are made when there have been three days since last collection and when there have been four days since last collection, and data were available on collection rates for these two conditions. In a three-times-per-week collection system, collections are made (twice) when there have been two days since last collection and (once) when there have been three days since last collection. Data for the latter condition were simply taken, unadjusted, from existing data; while histograms for the condition of two elapsed days since last collection were extrapolated. In any real study, it would be advisable to actually perform three-times-per-week collection in a few areas of different population densities and gather valid data for collection rates. It should also be noted that no allowance was made for the possibility of increased waste generation due to the increased frequency.

The results of these trials indicated that in order to collect three times per week, the number of collection vehicles operated within the area would need to be increased from 29 to 36. Amortization, maintenance, and operation (excluding labor) of these additional 7 trucks caused an increase in cost per ton collected of $0.58 while additional labor cost in the entire system amounted to $0.38 per ton. Thus, total cost increased from $10.68 per ton to $11.64 per ton, or about 9 percent.

There is, however, a possible additional cost which is not directly indicated by the model. If collection is performed twice per week, then on each of Monday, Tuesday, and Wednesday one is collecting four days' waste accumulation from one-third of the area or the equivalent of 1.33 days' accumulation over the entire area. If collection is performed three times per week, then on each of Monday and Tuesday one is collecting three days' waste accumulation from one-half of the area, or the equivalent of 1.5 days' accumulation over the entire area. Thus, the peak daily loads on transfer stations and disposal facilities increase by nearly 30 percent. There is, of course, an assumption that waste generation rates are the same on every day of the week. In actual practice (and in the data used by the model), this is not true; nonetheless, there is likely to be some increase in peak loads which might well require increased capital investment or increased overtime costs in the operation of facilities. Since the model does not include costs of operation of disposal facilities, and since there was no transfer station included in these runs, a direct indication of these costs was not provided; however, the model did show that peak daily loads at the disposal site increased from 297 tons per day to 341 tons per day, or approximately 15 percent.

Additional runs of the model were used to explore the potential savings associated with the installation of a transfer station within the collection area. Data on construction and operation costs of a transfer station were obtained from other cities, as were data on the speeds and costs of long-haul vehicles. A potential site for the location of a transfer station was selected and the coordinates of this site were provided to the model. In the first runs, the existing location of the incinerator (approximately 8 miles from the centroid of the collection area) was used. The results showed that for both two- and three-times-per-week collection, costs with and without the transfer station were approximately equal. The number of collection vehicles required with twice-per-week collection was reduced from 29 to 24; with three-times-per-week collection it was reduced from 36 to 30. However, the resulting reduction in collection cost was equalled by the cost of amortization and operation of the transfer station and the amortization, operation and labor costs of the three tractors and six to eight trailers required to haul to the disposal site.

Because of the lack of an indication of economies associated with a transfer station, the model was further exercised by relocating the disposal site both closer to and further from the collection area. It was found that, for twice-per-week collection, the cost per ton varied with distance from the collection area approximately according to the following linear equations:

a) Without transfer station:
Cost/ton = \$9.00 + \$0.20/mile
b) With transfer station:
Cost/ton = \$10.30 + \$0.03/mile

The intersection of these two lines is at approximately 8 miles; thus, for distances between collection area and disposal site of under 8 miles a transfer station in the collection area is uneconomical while for greater distances the transfer station rapidly develops significant savings. For three-times-per-week similar linear equations were found; their fixed costs were higher and their slopes were approximately the same as those shown above, and their intersection point was also at a distance of approximately 8 miles. These results cannot, of course, be used to develop any general conclusions about transfer stations; they are valid only for the particular area studied and the specific configuration of vehicles and other equipment which was assumed.

It is also possible to use the model to investigate the advantages of including a compaction unit in the transfer station. This is done by increasing the capital cost of the transfer station by the cost of the compaction unit and increasing the density of the waste in the long haul vehicle. Several runs were made using the model with a compaction unit designed to increase density in the long-haul trailers from about 450 pounds per cubic yard to about 800 pounds per cubic yard; the results indicated that savings associated with such a unit are not linear with transport distance and that rather large transport distances (in excess of about 15 miles) are necessary in order for the compaction unit to demonstrate any significant savings over the transfer station alone.

The Jacksonville Study (Reynolds, Smith, and Hills, Inc. [45]). A complete solid waste systems study was performed in 1970 in the City of Jacksonville, Florida by the engineering firm of Reynolds, Smith, and Hills. The need for this study was made apparent by the 1968 consolidation of Jacksonville and the surrounding Duval County, resulting in a patchwork of different solid waste systems. The entire study was unusually comprehensive, encompassing a complete fiscal and management review as well as recommendations for the immediate future and for long-range policy. It is beyond the scope of this chapter to discuss the entire study; attention is devoted herein to some uses of modelling in the decision process.

The Truitt simulator described above was adapted for use in a portion of the study. Since the consultants had available an IBM 1130 computer, the model was first converted to this machine and tested using the original Baltimore data. Modifications were then made to the model to permit consideration of either a 4-day or a 6-day workweek, since both of these were possibilities for the city. A great deal of difficulty was experienced by the consultants in making these and other minor modifications.

This emphasizes the problem associated with the construction of a general purpose simulator intended for use in many different cities under many different operating policies. Following this restructuring and extensive data collection, the model was used first to simulate current municipal operation of 33 trucks collecting twice weekly in a 4-day workweek. Cost of operation was found to be $13.46 per ton, which correlated closely with actual figures.

This model was then used to explore various comparatively short-run alternatives which were under consideration for the city. For example, the use of a transfer facility at an abandoned incinerator site was found to reduce the required number of vehicles to 27 but to produce significant delays in queueing at the transfer station resulting in a large amount of overtime. Addition of several more unloading bays in the transfer station reduced the overall cost to $11.90 per ton. A simulation using a backup transfer station site about 1.5 miles further away from the collection area produced similar results leading to the conclusion that the backup site was also economically feasible. Simulation of the use of a new disposal site without transfer station indicated that if the current disposal site were abandoned and the new one used, the required number of trucks would increase to 36 and the collection cost would increase to $14.78 per ton.

Because the consultants felt that a six-day workweek might prove to be more economical, several simulation runs were made using various combinations of workweek length, transfer stations, and disposal site. The results of these runs are summarized in the following table:

Collection Frequency	Workweek Length	Transfer Station	Number of Trucks	Total Cost ($/ton)
2/week	4 days	No	33	13.46
2/week	4 days	Yes	27	11.90
2/week	6 days	No	22	12.58
2/week	6 days	Yes	18	12.01
3/week	6 days	No	29	16.46
3/week	6 days	Yes	26	15.35

These results, coupled with other portions of the study, led to the recommendation that a collection frequency of twice per week in a six-day workweek should be used, if possible. Two additional runs of the simulator explored the feasibility of a proposed new regional landfill and a proposed new regional incinerator.

A facility location optimizing model (Skelly [50]), similar to that of Marks but including location of disposal facilities as variables, was also used in this study. The model was used principally to explore long range (to 1990) regional configurations of the entire system, and showed that the most economical arrangement consisted of two regional landfills with seven transfer stations.

IV. USE OF OPTIMIZING AND SIMULATION MODELS

The studies discussed above demonstrate that models are not generally sufficient to provide final answers to any management problem. Their greatest benefit may well be to assist decision makers in developing intuition and understanding the general behavior of a system. Thus, no single run of any model can be made to determine a solution; rather, many runs using different assumptions should be made.

Simulation models and optimizing models are inherently different. Optimization, because the solution technique requires particularly simple mathematical relationships, is usually done on a skeletal model which includes only the major components of the system and reflects only the major interrelationships between these components. Simulation can be done with a model containing much more detail but requires much more computer time and much more data. Further, simulation can only explore the consequences of a particular action or configuration, while a single run of an optimization model determines the appropriate configuration. As a result, perhaps the best use of the two kinds of models would be in series, applying optimization models to screen large numbers of alternatives and select a few which appear best, and then using simulation to explore in greater detail the consequences of these few alternatives. It is unfortunate that in the Jacksonville study the Truitt simulator was inadequate to handle the regional system which was explored by the Skelly model.

An additional consideration in the use of simulation is the current lack of general purpose simulators. This places each local area in the position of having to build its own simulator from scratch or modify an existing program to fit its own situation. Either of these is a formidable task. It is likely that a carefully constructed simulator can be modified as time goes on to reflect all changes in a system so that a city which made the initial investment required to build a simulator would always possess an up-to-date program but no successful efforts of this sort have yet been reported.

References

1. Altman, S., E. Beltrami, S. Rappaport, and G. Schoepfle, "A Nonlinear Programming Model of Crew Assignments for Household Refuse Collection," *IEEE Transactions on Systems, Man and Cybernetics*, Vol. SMC-1, No. 3, 1971.
2. Altman, S., N. Bhagat, and L. Bodin, "Extension of the Clarke and Wright Algorithm for Routing Garbage Trucks," presented at the VIII International Meeting of the Institute of Management Science, 1971.
3. Anderson, L. E. and A. K. Nigam, *Comprehensive Studies of Solid Wastes Management: A Mathematical Model for the Optimization of a Management System*, SERL Report 68-1, Sanitary Engineering Research Laboratory, University of California at Berkeley, 1968.
4. Andros, P. J., *Subdivision of Large Street Networks into Refuse Collection Districts*, M. S. Thesis, Clarkson College of Technology, 1972.
5. Beltrami, E. J. and L. D. Bodin, *Networks and Vehicle Routing for Municipal Waste Collection*, Urban Policy Sciences Report UPS 72-8, State University of New York at Stony Brook, 1972.
6. Bodin, L. D., "Towards a General Model for Manpower Scheduling," Parts I and II, *Jour. of Urban Analysis*, 1, No. 2, 1972.
7. Bodner, R. M., E. A. Cassell, and P. J. Andros, "Optimal Routing of Refuse Collection Vehicles," *Jour. San. Eng. Div., ASCE*, 96, No. SA4, 1970.
8. Brown, R. S., *An Analytical Model for District Assignment in Solid Waste Collection*, M. S. Thesis, Massachusetts Institute of Technology, 1972.
9. Clark, R. M. and B. P. Helms, "Decentralized Solid Waste Collection Facilities," *Jour. San. Eng. Div., ASCE*, 96, No. SA5, 1970.
10. Clark, R. M. and B. P. Helms, "Fleet Selection for Solid Waste Collection Systems," *Jour. San. Eng. Div., ASCE*, 98, No. SA1, 1972.
11. Clark, G. and J. W. Wright, "Scheduling of Vehicles from a Central Depot to a Number of Delivery Points," *Jour. Operations Research Society of America* 12, 1969.
12. Coyle, R. G. and M. J. C. Martin, "The Cost Minimization of Refuse Collection Operation," presented at the National Meeting of the Operations Research Society of America, 1968.
13. Dantzig, G. B. and J. H. Ramser, "The Truck Dispatching Problem," *Management Science* 6, No. 1, 1959.
14. Edmonds, J., "Maximum Matching and a Polyhedron with 0-1 Vertices," *Jour. of Research of National Bureau of Standards*, Sec. 69B, 1965.

15. Edmonds, J. and E. Johnson, "Matching, Euler Tours, and the Chinese Postman," *Mathematical Programming*, Vol. 5, No. 1, 1972.

16. Esmaili, H., "Facility Selection and Haul Optimization Model," *Jour. San. Eng. Div., ASCE, 98*, No. SA6, 1972.

17. Fuertes, L. A., J. F. Hudson, and D. H. Marks, *Analysis Models for Solid Waste Collection*, Report R72–70, Department of Civil Engineering, Massachusetts Institute of Technology, 1972.

18. Golueke, C. G., *Comprehensive Studies of Solid Wastes Management: Third Annual Report*, SERL Report 70–2, Sanitary Engineering Research Laboratory, University of California at Berkeley, 1970.

19. Golueke, C. G. and P. H. McGauhey, *Comprehensive Studies of Solid Wastes Management: First Annual Report*, SERL Report 67–7, Sanitary Engineering Research Laboratory, University of California at Berkeley, 1967.

20. Golueke, C. G. and P. H. McGauhey, *Comprehensive Studies of Solid Wastes Management: Second Annual Report*, SERL Report 69–1, Sanitary Engineering Research Laboratory, University of California at Berkeley, 1969.

21. Helferich, O. K., V. Hoffner, and D. E. Gee, "Dynamic Simulation Model for Planning Solid Waste Management," presented at XIX International Meeting of the Institute of Management Sciences, 1972.

22. Helms, B. P. and R. M. Clark, "Locational Models for Solid Waste Management," *Jour. Urban Planning and Development Div. ASCE, 97*, No. UP1, 1971.

23. Klee, A. J., "Let DARE Make Your Solid Waste Decisions," *The American City, 85*, No. 2, 1970.

24. Klee, A. J., "DISCUS—A Solid Waste Management Game," *IEEE Transactions on Geoscience and Electronics*, Vol. GE–8, No. 3, 1970.

25. Klee, A. J., "The Role of Decision Models in the Evaluation of Competing Environmental Health Alternatives," *Management Science*, 18, No. 2, 1971.

26. Liebling, T. M., *Graphentheorie in Planungs—und Tourenproblemen am Beispiel des Stadtischen Stranssendienstes*, Volume 21 of Lecture Notes in Operations Research and Mathematical Economics: Springer-Verlag, New York, 1970.

27. Liebman, J. C., "Mathematical Models of Solid Waste Collection and Disposal," presented at 38th National Meeting of the Operations Research Society of America, 1970.

28. Liebman, J. C., "Systems Approaches to Solid Waste Collection Problems," Chapter 15 in *Systems Approach to Environmental Pollution*, G. K. Chacko (ed.), Operations Research Society of America, Arlington, Va., 1972.

29. Lofy, R. J., *Techniques for Optimal Routing and Scheduling of Refuse Collection Vehicles*, Ph.D. thesis, University of Wisconsin, 1971.

30. Male, James W., *A Heuristic Solution to the M-Postmen's Problem*, Ph.D. Thesis, The Johns Hopkins University, 1973.

31. Marks, D. H., J. Cohon, H. Moore, and R. Stricker, "Routing for Municipal Services," *Proceedings*, 6th Annual ACM Urban Symposium, 1971.

32. Marks, D. H. and J. C. Liebman, *Mathematical Analysis of Solid Waste Collection*, Public Health Service Publication 2104, U. S. Government Printing Office, 1970.

33. Marks, D. H. and J. C. Liebman, "Optimal Transfer Facility Location for Solid Waste Collection," presented at 37th National Meeting of the Operations Research Society of America, 1970.

34. Marks, D. H. and J. C. Liebman, "Location Models: Solid Waste Collection Example," *Jour. Urban Planning and Development Div. ASCE, 97*, No. UP1, 1971.

35. Marks, D. H. and R. Stricker, "Routing for Public Service Vehicles,' *Jour. Urban Planning and Development Div. ASCE, 97*, No. UP2, 1971.

36. Martin, M. J. C. and A. E. Rothwell, "Cost and Manpower Minimization of Refuse Collection Operations: Some Applications of a Basic Model," presented at 39th National Meeting of the Operations Research Society of America, 1971.

37. Morse, N. and E. W. Roth, *Systems Analysis of Regional Solid Waste Handling*, Public Health Service Publication 2065, U. S. Government Printing Office, 1970.

38. Nigam, A. K., *Comprehensive Studies of Solid Wastes Management: Optimal Strategies in Capacity Expansion*, SERL Report 70–10, Sanitary Engineering Research Laboratory, University of California at Berkeley, 1970.

39. Owen, E. H., "Computer Program Cuts Cost of Urban Solid Waste Collection," *Public Works*, January 1970.

40. Quon, J. E., A. Charnes, and S. J. Wersan, "Simulation and Analyses of A Refuse Collection System," *Jour. San. Eng. Div., ASCE, 91*, No. SA5, 1965.

41. Quon, J. E., M. Tanaka, and A. Charnes, "Refuse Quantities and the Frequency of Service," *Jour. San. Eng. Div., ASCE, 94*, No. SA2, 1968.

42. Quon, J. E., M. Tanaka, and S. J. Wersan, "Simulation Model of Refuse Collection Policies," *Jour. San. Eng. Div., ASCE, 95*, No. SA3, 1969.

43. Rao, H. S., M. H. Richardson, and D. A. Wismer, "Systems Analysis in Solid Waste Management Planning," presented at Joint National Conference on Major Systems, Anaheim, California, 1971.

44. ReVelle, C. S., D. H. Marks, and J. C. Liebman, "An Analysis of Private and Public Sector Location Models," *Management Science, 16*, No. 11, 1970.

45. Reynolds, Smith, and Hills, Inc., *Final Engineering and Management Report, Solid Waste Systems Study, City of Jacksonville, Fla.* 1970.

46. Rossman, L. A., *A General Model for Solid Waste Management Facility Selection*, Special Project, Department of Civil Engineering, University of Illinois at Urbana-Champaign, 1971.

47. Rothgeb, W. L., "Computerized Refuse Collection," *Public Works, 101*, No. 4, 1970.

48. Schultz, G. P., *Managerial Decision Making in Local Government: Facility Planning for Solid Waste Collection*, Ph.D. Thesis, Cornell University, 1968.

49. Shuster, K. A. and D. A. Schur, *A Heuristic Approach to Routing Solid Waste Collection Vehicles*, U.S. Environmental Protection Agency, Office of Solid Waste Management Programs, 1972.

50. Skelly, M. J., *Planning of Regional Refuse Systems*, Ph.D. Thesis, Cornell University, 1968.

51. Stern, H. I., *Comprehensive Studies of Solid Wastes Management: Optimal Service Policies for Solid Waste Treatment Facilities*, SERL Report 69-6, Sanitary Engineering Research Laboratory, University of California at Berkeley, 1969.

52. Stricker, R., *Public Sector Vehicle Routing: The Chinese Postman Problem*, M.S. Thesis, Massachusetts Institute of Technology, 1970.

53. Tanaka, M. and J. E. Quon, "A Linear Programming Model for the Selection of Refuse Collection Schedules," presented at the 40th National Meeting of the Operations Research Society of America, 1971.

54. Truitt, M. M., J. C. Liebman, and C. W. Kruse, "Simulation Model of Urban Refuse Collection," *Jour. San. Eng. Div., ASCE, 95*, No. SA2, 1969.

55. Truitt, M. M., J. C. Liebman, and C. W. Kruse, *Mathematical Modeling of Solid Waste Collection Policies*, Public Health Service Publication 2030, U.S. Government Printing Office, 1970.

56. Wahi, P. N. and T. I. Peterson, "Management Science and Gaming in Waste Management," *Jour. San. Eng. Div., ASCE, 98*, No. SA5, 1972.

57. Walker, W. L., "A Heuristic Adjacent Extreme Point Algorithm for the Fixed Charge Problem," New York City-Rand Institute Report P-5042, June, 1973.

58. Wathne, M., *Optimal Routing of Solid Waste Collection Vehicles*, Ph.D. Thesis, The Johns Hopkins University, 1972.

59. Weber, A., *Uber den Standort der Industrien*, Tubingen, 1909. Translated as Friedrich, C. J., *Alfred Weber's Theory of Location of Industries*, Chicago, 1929.

60. Wolfe, H. and R. Zinn, "Systems Analysis of Solid Waste Disposal Problems," *Public Works, 98*, September 1967.

61. Wyskida, R. M. and J. N. D. Gupta, "A Routing Problem in City Solid Waste Collection," presented at the 39th National Meeting of the Operations Research Society of America, 1971.

Chapter 6

Models in Urban Development

By

James C. Ohls and Peter Hutchinson
(with contributions by)
T. R. Lakshmanan and Philip D. Patterson

SUMMARY 167

I. INTRODUCTION 167

II. SOME GENERAL THEMES IN URBAN MODELING 169

Large Number of Variables 169
Large Number of Policy Parameters 169
Variables and Parameters Highly Intercorrelated 170
Long Time Horizon 170
Data Limitations 170
Objectives of Urban Models 171
Simplifying Techniques 172
Evaluation Criteria 172

III. MAJOR REVIEWS OF URBAN DEVELOPMENT MODELS 173

Introduction 173
Twelve Selected Surveys 174
European Surveys 179

IV. SELECTED URBAN DEVELOPMENT MODELS 179

Partial Analytic Models 181
NBER Model—The Residential Sector 181
The Market Potential Model—The Retail Sector 184
Comprehensive Analytical Models 185
The Lowry Model 185
The Urban Dynamics Model 190
Simulation-Gaming Models 193
CLUG—Community Land Use Game 193
RIVER BASIN MODEL 193
APEX—Air Pollution Exercise 195

V. CONCLUDING COMMENTS 196

REFERENCES 198

Models In Urban Development

SUMMARY

Public officials at both the local and Federal levels must make important decisions affecting urban structure and urban growth. Zoning, urban renewal, poverty programs, transportation planning, and housing programs are just a few of the many areas of public policy which are influenced by and have effects on urban spatial location. Because of the great complexity of urban systems, modeling may have great potential as an aid to formulating and analyzing public programs in these and other urban policy areas.

A substantial number of models of urban structure have been developed, and they vary considerably with regard to general approach and to specific details. Urban models of necessity involve simplification in order to make them manageable, and the ways in which specific modelers have chosen to accomplish this simplification can best be understood in relation to the objectives which the various models have been designed to accomplish.

The government funding for urban models dealing with predicting the size and distribution of land using activities within a metropolitan area has been so widespread over the past twenty years, that a substantial literature has already been developed comparing the models and assessing the value of efforts made to date. This body of literature has received recent inputs from a number of researchers who are critical of the past modeling efforts. However, the funding of urban models continues as the model builders reduce their claims and expand their methodologies.

This chapter provides an overview of the general themes in urban modeling and of the major urban modeling efforts during the past two decades by first reviewing some previous surveys of urban models. Then several selected urban development models are examined in some detail. The chapter concludes with a discussion of recent critiques of urban development modeling and a view of the policy role of urban development models.

The summary table shows the urban models dealt with in this chapter.

I. INTRODUCTION

Urban areas are highly complex forms of economic and social organization. Almost by definition of the term, a city involves large numbers of people and large amounts of economic resources, all located in close physical proximity to one another and all engaged in complex patterns of interaction. Public sector decision making in the urban context—if it is to be effective in accomplishing policy objectives—must be characterized by as great an understanding as possible of the complex interactions between different urban variables. Urban policy makers must be aware not only of the immediate effects of their decisions but of the indirect effects as well. The urban policy maker who authorizes the construction of a new subway line, for instance, must have as sound an understanding as possible not only of the immediate effects of the line in speeding travel for existing commuters, but also of such indirect effects of the new line as changed industrial location patterns, changed job opportunities for minority groups, changed tax revenues,

Preceding page blank

URBAN DEVELOPMENT MODELS

Models Discussed in Chapter

Summary Table

Model	General Type	Important Characteristics
NBER	Analytic, Partial	Combines empirical and theoretical approaches in a dynamic model for housing.
Retail Market Potential	Analytic, Partial	Locates large retail trade centers using drawing power for customers. Is intimately tied to a transportation model.
Lowry and **PLUM**	Analytic, Comprehensive	Industrial land use is input exogenously and housing and retail locations are derived. Several levels of refinement and sophistication.
Urban Dynamics	System Dynamics, Comprehensive	Simulates growth or industry, residences, and population within a fixed area over a very long time period (50–250 years).
CLUG	Simulation-Gaming	Players make locational decisions. Payoffs based on location theory concepts for housing, retail and industry.
RIVER BASIN	Simulation-Gaming	Players make locational and other decisions as economic, social, and governmental decision makers. Computer simulation of population movements and market operations for housing, retail and industry.
APEX	Simulation-Gaming	Players make some of the locational decisions and computer model simulates population movements, business operations, and voting behavior for housing and industry.

and changed needs for other public services, all of which may result from his decision.

It is no accident, therefore, that urban model building has received a great deal of attention in the past two decades. Modeling techniques, which allow their user to analyze complex, simultaneous relationships, would appear to be especially well-suited to dealing with urban phenomena and indeed there has been great interest recently in constructing urban models and in using them in policy analysis. Spurred both by increased concern about urban problems and by recent advances in modeling techniques, researchers and policy analysts have constructed literally hundreds of such models in the past two decades, and many have been formally published. They have sought to accomplish a broad range of specific objectives, from understanding the underlying determinants of spatial location within the city, to analyzing the causes of city growth and decay, to predicting future land uses in specific parts of urban areas. The models' developers have come from a broad range of backgrounds in both the social and natural sciences. This chapter will describe some previous surveys of urban models, survey some of the most important advances which have been made in constructing urban models and show how these models have been applied to policy making.

The term "urban model" must be defined with some care. Taken broadly, it might well apply to a high proportion of all the models described not only in this chapter but in this entire report. Since the United States is a highly urbanized society, most policy making in this country takes place in an urban environment. Hence, many of the models described in other chapters of the report are set in an urban context and deal with certain subsectors of urban phenomena. For the purposes of this chapter, however, we shall adopt a more narrow definition of urban model. Specifically, we shall limit our attention only to models which attempt to describe land use—housing, commercial and industrial—and urban development processes within an entire metropolitan area. Thus models of, for instance, educational planning or criminal justice which are not concerned with urban development will not be discussed. Not all of the models which we shall consider deal in an equally comprehensive way with all facets of urban development. For reasons of simplicity, certain subsectors of the urban process are suppressed in some of the models and emphasized in others.

II. SOME GENERAL THEMES IN URBAN MODELING

A clear conception of the high degree of complexity and interrelatedness of activity in an urban area is crucially important in understanding the form in which various urban models have been developed. In this section of the paper we shall explicitly examine the reasons for this complexity. Doing so will allow us to better understand the complexity itself, and it will also provide a convenient context for discussing certain elements of urban development and urban policy making which will be referred to in our later discussions of specific models. Four general reasons for the great complexity of urban systems will be discussed below. They are (1) the large number of relevant variables, (2) the large number of policy parameters available to the urban public decision maker, (3) the fact that variables and parameters are highly intercorrelated, and (4) the extremely long time horizon facing the urban decision maker.

Large Number of Variables

One of the reasons for the high degree of complexity of urban phenomena is the large number of relevant variables which must be included in any reasonably complete discussion of urban development. Housing, commercial activity, industrial activity, natural resources, transportation, and public services are just a few of the urban variables of interest to policy makers, and these must be considered as they relate to literally millions of people and thousands of businesses.

Large Number of Policy Parameters

Another factor which contributes to the complexity of the task facing the urban model builder who seeks to construct a model which is useful in policy making is that, not only is there a large number of key variables, but there are also a large number of policy parameters and policy-making contexts in which models may be useful.

At the local level, for instance, government officials make many different kinds of decisions which impinge on the urban development process. For instance, zoning decisions determine what kinds of economic activity will be allowed to locate in various parts of the urban area. Transportation planning—both the choice between alternative modes and the choice of specific locations, schedules, and prices within modes—also provides wide scope for local decision making. Poverty programs and programs aimed at aiding disadvantaged minorities to obtain greater economic opportunity also provide important policy questions for local officials. Local decision makers also deal with a broad range of local housing programs including public housing and the enforcement of building codes.

There is also a need for an understanding of the development process and for the use of urban models at the Federal level. Many of the local programs listed above, for instance, include Federal participation and hence involve Federal as well as local decision makers. Furthermore, there are many programs conducted principally at the Federal level which seek to have an impact largely in urban areas. Federal housing subsidy programs are an example of these. Another general policy area where Federal policy makers must be concerned with urban phenomena is the general revenue sharing program which has recently been enacted. This program was largely developed in response to fiscal problems perceived by governments in urban areas, and urban models have been developed which attempt to provide insight into the role of local government finances in influencing the urban development process.

Variables and Parameters Highly Intercorrelated

Thus we have seen that urban systems are made complex by the existence of both large numbers of key variables and large numbers of policy parameters in the urban context. A third element in the complexity of the urban development process is the fact that all of the variables and policy parameters are highly interconnected. Large numbers of variables and policy parameters need not, in themselves, lead to a complex system. If it were possible to separate them into largely separate subsectors, the urban system could still be modeled relatively simply. In fact, however, all of the variables in the system are highly related to one another. For instance, a change in the transportation system is likely to have important impacts on housing and industrial location which in turn are likely to affect the need for provision of public services which in turn may affect the ability of the government to provide transportation services, etc. Or, to take one more example, a public housing program may affect the number of potential workers in a given part of the city, which may influence industrial location, which may have an impact on transportation planning, which may have a further impact on the location of poor people and on public housing decisions, etc.

Long Time Horizon

Besides the large number of important variables and parameters and their interrelatedness, there is still one other important factor which should be mentioned as contributing to the complexity of the urban model builder's task. This factor is the need for an extremely long time horizon in making many urban policy decisions. This need arises because of the extremely long-lived nature of residential, industrial, and other types of investment in cities. Most structures last at least 50 years, and some last for centuries. This means that policy decisions made today will have important and to some extent irreversible impacts on development processes for decades to come, and this adds to the complexity of the model builder's task because he must attempt to construct a model which is relevant in gaining insight into events happening not only at the present time but many years into the future.

Data Limitations

Summarizing, then, we have seen that an important problem facing the urban modeler is the complexity of the relationships which he is attempting to represent with his model. We shall see below that many techniques used in urban models can be best understood as attempts to handle this problem of complexity. Before discussing this, however, it will be useful to consider a second general problem faced by the urban model maker. This second problem is data limitations. Accurate information on many key variables in the urban system are often simply not available. For instance, except for decennial years, very little information is available concerning even such a basic variable as the total population in a given city. And even in decennial years, when the U. S. Census of Population and Housing is undertaken, there is substantial evidence that the data collected is not highly reliable—especially in providing information about minority groups. Similarly, detailed data concerning the nature of industrial activity in a city is typically only available in the years when a U. S. Census of Manufacturing is conducted.

Furthermore, even when data is collected for a key variable in a city, it is often not available in a usable form. For instance, in many cities property tax and deed transfer records provide substantial amounts of data concerning individual parcels of land. But this data is often not assembled in such a way as to provide systematic samples which could be useful in research and model building.

There are a number of important reasons for the unavailability of data. One of them is that most of the key parameters for an urban model are behavioral in nature. For instance, it is crucial to have information about such behavioral questions as what influences families' and firms' locational decisions or what the demand for trips on a new transport system will be. And because these parameters are behavioral rather than technological in nature, their quantification is conceptually difficult and, in practice, imprecise.

Furthermore, in most cases data gathering through experimentation is prohibitively expensive. A natural scientist attempting to model a chemical reaction can gain data by repeatedly undertaking and observing experiments with the chemicals, but such a procedure is clearly not possible with regard to gaining information about the results of a multimillion dollar transportation investment. For the most part the urban modeler must rely on such information as the accidents of history provide him with. (As a qualification to this, incidentally, it should be pointed out that there has been increased interest in recent years in forms of social experimentation. For instance, the U. S. government is currently undertaking important experimental type programs testing alternative forms of the negative income tax and alternative forms of housing subsidies.)

Still another practical problem in assembling data which should be mentioned results from the fact that there are often many governmental jurisdictions within an urban area, and these different jurisdictions are often not consistent in the way in which they gather and present the data which they collect.

Objectives of Urban Models

We have seen, then, that both the complexity of the urban system and data limitations create difficulties for urban modelers. In order to gain further insight into the exact nature of these difficulties and into the ways in which model builders have sought to overcome them, we must turn now to a brief discussion of the various specific objectives of urban modelers. Of course, the list of exact objectives is as long as the list of urban models themselves, but it is useful to distinguish four general kinds of objectives which seem to characterize most of the models which have been developed. The four objectives which we shall discuss are (1) prediction, (2) policy analysis, (3) optimization, and (4) gaining a better fundamental understanding of the nature of urban processes.

In principal, of course, a single model could be developed which accomplished all four of the above objectives. If one could create a model which provided a full representation of the urban development process, it could be used for prediction, it would contain within it key policy parameters for policy analysis, it would allow one to optimize any given objective function, and it would —by its basic structure—give insight into the basic forces underlying observed urban phenomena. Unfortunately, for the reasons discussed above, such an all-purpose model is probably not feasible at the present time. Given current limitations in modeling and solution techniques, both the complexity of urban phenomena and also problems concerning the available data make it extremely difficult to construct one single urban model which can completely satisfy all four of these objectives. As a result of these difficulties, therefore, the urban modeler is forced to make compromises between the various objectives and to adopt simplifying techniques which will make his task manageable, even at the price of lost completeness. The simplifying techniques chosen depend, of course, on the principal objectives which a given model seeks to accomplish, and, as an aid to understanding the individual models to be discussed below, it will be useful to examine a number of general approaches which have been used in attempting to simplify urban models to the point where they can be made manageable in the context of present-day modeling techniques. Three simplifying techniques will be considered: (1) aggregation, (2) supressing certain sectors of activity in the model, and (3) taking some sectors of activity as exogenous.

Simplifying Techniques

One frequently-chosen simplifying technique is aggregation. This can take a number of forms. For example, many models rely on what we shall call sectoral aggregation. Certain sectors of the urban economy are simply aggregated together. For instance all forms of industrial and commercial activity may be treated as a single kind of activity.

Another form of aggregation may be geographical aggregation. Instead of treating each parcel of land or each small area of a city as an individual unit, modelers have often found it convenient to lump geographical subunits together into larger groups. For instance, all parcels of land in the urban area may be divided into two groups, those in the center city and those in the suburbs. Or, in some cases, geographical distinctions may be ignored entirely and land in the entire area may be taken as spatially homogeneous.

A second general technique for simplifying urban models is to entirely suppress certain sectors of activity in the model. For instance, in a model concerned only with residential location, the entire industrial sector may be ignored. Or in a model concerned with the economic interactions between households and firms, the existence of minority groups may be ignored.

Finally, a third technique for simplifying models is to take some sectors as exogenous. A model may, for instance, include the spatial location of firms but assume that the firms' locations are predetermined outside of the context of the variables which are allowed to vary in the model. Or, to take another example, transportation prices may be included in a model but their magnitude and the location of alternative transport modes may be taken as given and not allowed to vary.

So far, we have pointed out that complete models of the entire complexity of the urban development process are very difficult to create given the current state of the modeling art, and we have discussed a number of techniques used to simplify urban models enough to make them manageable. We shall now conclude these general remarks about urban modeling by attempting to identify criteria by which various urban models can be evaluated.

Evaluation Criteria

It is probably not appropriate to judge models by their overall correspondence to reality. Inevitably, any model so evaluated will be found to depart from reality in many ways. Indeed the essence of modeling is abstracting from some facets of reality in order to focus attention on other aspects. Hence more limited criteria are necessary. In the following paragraphs we shall identify four criteria which can be useful in appraising and comparing urban models. These four criteria are (1) inclusion of key variables, (2) simplicity, (3) consistency with theory, and (4) consistency with data.

The first criterion is whether or not the modeler has included all of the variables which are necessary to accomplish the objectives of his model. For instance a model which is to be useful for making specific zoning decisions must include a good deal of geographical detail. To engage in significant geographical aggregation would defeat the model's purpose. On the other hand, a model constructed to aid planning of financial aid to the cities by the Federal government might appropriately make use of a great deal of geographical aggregation. Or, to take another example of this criterion, a model which was aimed at providing a useful planning tool for local planners could perhaps take *Federal* housing policy as exogenous. But a model which is to be used as a planning tool by the Federal Department of Housing and Urban Development must clearly include Federal housing policy as a variable policy parameter.

Simplicity provides a second criterion with which to appraise urban models. It is the very essence of modeling that it is often useful to focus attention on key variables in a system so that they can be understood adequately, and, as emphasized in our discussion of the first criterion above, important variables certainly must be included in a model. But it may often be counterproductive to clutter a model with a large number of variables which are not crucial to the model's basic purpose. Inclusion of these extraneous variables may make the model so complicated that the reader is unable to gain any true understanding of the basic relationships between the key parameters in which he is interested.

A third criterion with which we shall evaluate the models discussed below is consistency with theory. Long traditions of research in a variety of disciplines including economics, political science, and urban planning provide the prospective urban model builder with a variety of theoretical tools to help him with his task. In light of this, it seems appropriate to examine the consistency of models with theory. For instance, there is a rich theory in economics which predicts that as the price of land goes up, all other things being held constant, families will react by consuming less land as part of their housing consumption. There is some presumption, therefore, that a model which deals with these variables but does not include this relationship may not be an adequate representation of reality. There may, of course, sometimes be reasons for ignoring such a known theoretical relationship. For instance, the modeler may have been aware of the theory but may have determined to his satisfaction that the relationship was not sufficiently quantitatively important to warrant its inclusion in his model. But in appraising urban models it is often useful to at least investigate the degree to which the model builder has taken account of existing theory.

Finally, another criterion which can be applied in evaluating urban models is their consistency with data. To be sure, data limitations such as those discussed above may often make rigorous attempts to fit a model to data impossible. And in any case, since all models necessarily involve abstracting from reality, perfect correspondence between variables in the model and real world data is not to be expected. Nevertheless, it is reasonable and often useful to ask whether or not the urban model builder, in constructing his model, has fully made use of whatever empirical resources were available. For a given model, at a given degree of aggregation, as much consistency as possible with the available data would appear to be desirable.

This completes our general remarks about urban modeling. We have discussed the basic variables which urban models seek to include; we have considered a number of difficulties inherent in urban model building; and we have identified a number of partial criteria with which urban models may be evaluated.

III. MAJOR REVIEWS OF URBAN DEVELOPMENT MODELS

Introduction

Many dollars have been spent annually during the past two decades on the design and testing of urban development models. This relatively long history of modeling of urban growth and change has generated a significant literature of reviews and surveys that this chapter will not attempt to duplicate. Instead, this section will describe some of the past survey efforts so that the reader may gain an appreciation of the scope of such activity and also so that the reader will be informed as to the best sources for more detailed model descriptions and comparisons.

Figure 1 shows a matrix consisting of twelve rows (selected model surveys) and 21 columns (selected urban development models). An "X" in the matrix indicates that the survey dealt with the corresponding urban model. The model surveys are listed in chronological order, but it must be borne in mind that the dates attached to the models refer to dates on which published material was available. These model dates may actually have preceded or followed the actual implementation of the models.

Before briefly discussing each survey, it might be useful to state the type of impetus behind each survey: "premodel requirement" or "academic." Almost half of the surveys were required as an initial step to a new (or contemplated) urban modeling effort. The Traffic Research Corporation survey preceded the development of the EMPIRIC model in Boston. The Donald Lamb effort was prepared for the Southwestern Pennsylvania Regional Planning Commission prior to a model construction effort there. Douglass Lee prepared his survey for the Cornell Aeronautical Laboratory and expanded upon it to produce a dissertation. Thus a joint commercial-academic product was achieved. The National Bureau of Standards conducted its survey on behalf of the Model Cities office of the Department of Housing and Urban Development. The purpose was to determine the applicability of decision-making models for the Model Cities programs at the city level. The Nagelberg and Little survey was conducted as part of the Institute for the Future's plan to develop an urban research laboratory.

The other surveys were stimulated mostly by academic forces. The May issue of the American Institute of Planners journal in 1965 was devoted entirely to articles dealing with urban development models. The journal provided this special issue for the educational benefit of its readers who are professional planners concerned with the urban development process. Ira Lowry prepared his classic article for original presentation at the June 1967 Conference on Urban Development Models conducted by the Highway Research Board. Britton Harris prepared his paper for original presentation at the Resources for the Future sponsored Conference on Urban Economics in 1967.

The book by Kilbridge et al. evolved out of the Urban Analysis Project sponsored by the Harvard Graduate School of Business during 1967–1969. The paper by George Hemmens was presented at a computer simulation techniques seminar series at George Washington University in 1970. The book edited by David Sweet grew out of a Models of Urban Structure Conference sponsored jointly by Battelle of Columbus and Ohio State University. The review article by Goldstein and Moses appeared in the June 1973 volume of the Journal of Economic Literature. It is one of the few surveys on urban models prepared by economists for economists.

The 12 surveys selected for brief comment in this chapter are not exhaustive but they are representative of the range of surveys made over the past decade. These twelve surveys are described briefly in the following section and then reference is made to several European surveys that describe urban development modeling abroad.

Twelve Selected Surveys

The Traffic Research Corporation Survey

This survey conducted in 1963 had very few urban models to describe since not many had been operated for specific geographic locations at that early date. Three of the four operational models listed by the TRC are found in the matrix in Figure 1. It is interesting to note that traces of these three models can be found in the EM-

	Urban Models																					
Surveys	CATS—Chicago Transportation Study (Chicago, 1960)	Pittsburgh Urban Renewal Simulation Model (Steger, 1965)	a. Industrial Location Model (Putman, 1963)	b. Lowry Model (Lowry, 1964)	c. TOMM—Time-Oriented Metropolitan Model (Crecine, 1964)	San Francisco Community Renewal Model (Arthur D. Little, 1966)	Penn-Jersey Models	a. Regional Growth (Almendinger, 1961)	b. Activities Model (Seidman, 1964)	c. TOMM—Time-Oriented Model (Crecine, 1964)	EMPIRIC (Hill, 1965)	POLIMERIC (Hill, 1965)	SEWRPC—Southeastern Wisconsin Regional Planning Commission (Schlager, 1965)	UNC—Univ. of North Carolina (Chapin, 1965)	Opportunity Accessibility Model (Lathrop, 1965)	Retail Market Potential Model (Lakshmanan, 1965)	BASS—Bay Area Simulation Study (Creve, 1968)	Land Use Game (Feldt, 1972)	CITY RIVER BASIN MODEL (Environmetrics, 1970b, EPA, 1972a)	METRO-APEX (EPA, 1972b)	URBAN DYNAMICS (Forrester, 1969)	NBER—National Bureau of Economic Research Model (Ingram, 1972)
1. Traffic Research Corporation (1963)	X					X			X					X								
2. AIP (Harris, 1965)		X				X					X		X	X	X	X						
3. Donald Lamb (1967)			X	X	X	X		X			X		X		X	X						
4. Douglass Lee (1968)			X	X	X	X	X	X	X		X		X	X								
5. Ira Lowry (1968)	X			X		X			X		X	X		X								
6. Britton Harris (1968)	X		X	X		X			X		X			X		X						
7. Maurice Kilbridge, Robert O'Block, Paul Teplitz (1970)	X	X		X	X	X	X	X			X	X	X	X	X	X						
8. George Hemmens (1970)				X					X		X											
9. National Bureau of Standards (Miller, 1970)								X									X	X	X	X		
10. Mark Nagelberg and Dennis Little (1970)	X	X		X		X			X		X	X		X				X	X	X	X	
11. David Sweet (1972)					X				X		X	X					X	X	X	X	X	
12. Gerald Goldstein and Leon Moses (1973)				X																		X

Figure 1—*Matrix of Urban Development Models and Urban Model Surveys*

PIRIC model that eventually evolved out of the TRC (now Peat, Marwick and Mitchell) effort. The TRC survey described five research-oriented techniques and five conceptual techniques in addition to the four operational urban development models. A brief evaluation and critique followed a description of the fourteen land use forecasting techniques contained in the survey.

The AIP Special Issue

In 1959, and again in 1965, the entire May issue of the Journal of the American Institute of Planners was devoted to urban models. The 1959 issue appeared too early to describe any of the urban models contained in the matrix. The 1965 issue contained an introduction and short glossary of terms by Britton Harris, the guest editor of the issue. In addition to articles by the designers of seven of the models listed in Figure 1, there was a future-oriented article by Bill Garrison describing what urban transportation planning models might look like ten years later. The major breakthrough expected by Garrison was in data handling and analysis.

Another excellent article in this AIP issue was the one by Ira Lowry entitled "A Short Course in Model Design." In this concise review article, Lowry discusses the uses of models (descriptive, predictive and planning), levels of aggregation, handling of time, concepts of change (distinction between stocks and flows), solution methods (analytic, iterative, machine simulation, and man-machine simulation), fitting or calibrating the model (variables and parameters), and testing the model (sensitivity and performance). Since a good portion of the literature with respect to urban models has applicability and relevance to other functional areas of modeling, this article by Lowry is a good example of a prototype article that could be adapted to other fields of modeling.

Donald Lamb Survey

Donald Lamb produced an informative review of existing land use models as a preliminary step in the development of a model for the Southwestern Pennsylvania Regional Planning Commission. His most telling criticism was that none of the ten major models he investigated were "concerned with the validity of the projected change in the regional level." That is, all of the models that he reviewed allocated growth that had been forecast exogenously by some technique or process independent of the models themselves. This criticism still applies to most urban development models.

More carefully than in most reviews, Lamb concentrated on answering the question: "How does the model work?" and "What does it mean analytically?" Lamb defined his terms carefully, presented clear and helpful examples of mathematical techniques (trend models and logic models), and provided a general framework for reviewing the ten models in his survey. For each model, Lamb provided a general description, presented its characteristics (including equations) and data requirements and showed the operational sequence.

Douglass Lee Survey

The research report by Douglass Lee (1968) entitled "Models and Techniques for Urban Planning" brought together some of the best description material on urban models, synthesized it, added to it, and thereby resulted in the single most useful secondary source of urban model information.

Lee discussed some of the dimensions of modeling (comprehensiveness, disaggregation, treatment of time, abstraction, and descriptive-behavioral-normative) that Britton Harris elaborated upon more thoroughly elsewhere (Harris, 1968). He summarized the paradigm developed by Ira Lowry (1967), and abstracted the other urban model typologies developed by Steger (1965), Traffic Research Corporation (1963), Donald Lamb (1967) and Kilbridge et al. (1968).

Lee also discussed the theories (such as the general equilibrium theory and the gravity model), methods (simulation, social accounting, and econometric modeling), and techniques (such as input-output, mathematical programming and mathematics). As part of his chapter on model construction, Lee clarified the distinction among components, variables (exogenous and endogenous) and relationships (functions and parameters). This chapter concentrated heavily on the substantial data demands of urban models and the importance of model objectives.

The Lowry Paradigm and Survey

Ira Lowry developed an urban land market paradigm that is a very useful way of looking at how urban models usually emphasized either the use to which a piece of land is put or the allocation of a particular activity to a spatial location in the metropolitan area. These two actions might appear to be identical but Lowry showed the subtle differences and used the paradigm to characterize the seven urban models he surveyed. This paper by Lowry is a classic in the field of urban modeling and can provide the reader with insights not easily gained in any other way. The seven types of models described by Lowry are (1) land use (CATS), (2) land use succession (land use over time-UNC model), (3) location (EMPIRIC), (4) migration (location over time-Polimetric), (5) hybrid (his own Lowry Model), (6) market demand (Penn-Jersey Model), and (7) market supply (the San Francisco Model).

Harris Paper

Like Lee and Lowry, Harris presents more than a survey of urban models. Rather than dealing with the many model issues discussed by Lee or adopting some paradigm approach as employed by Lowry, Harris concentrated his attention on six fundamental dimensions of (urban) models:

1. Descriptive vs. Analytic
2. Holistic vs. Partial
3. Macro vs. Micro
4. Static vs. Dynamic
5. Deterministic vs. Probabilistic
6. Simultaneous vs. Sequential

He then went on to show how these dimensions have been applied in the case of retail location models, residential location, manufacturing location, and the location of other types of land uses (such as transportation facilities, open space, office space, etc.). This article by Harris is quoted often in the urban modeling literature and his dimension scheme has been used in other types of modeling as well.

Book by Kilbridge et. al.

This book devoted only three of its eleven chapters to previously built urban models, and two of these three were devoted almost exclusively to the abstraction, verification, validation and application of the TOMM model. The remaining chapters dealt with subjects such as the role of models in urban planning, population density concepts and measures, and economic models for housing analysis.

The opening chapter classified 20 well known models according to subject (land use, population, transportation, and economic activity), function (projection, allocation and deviation), theory (behavioral, gravity, trend, and input-output), and method (econometric, mathematical programming and simulation). This classification framework is nice and neat but the simplification necessary to categorize a complete model according to such a rigid structure has its pitfalls.

Hemmens Seminar Paper

In 1970, George Hemmens, who edited the Highway Research Board's special entitled "Urban Development Models," made some observations about the state of urban development modeling. To him, the models make land use highly dependent upon accessibility (illustrating the problem of making this chapter independent of the chapter on Transportation Models), deal with highly integrated socio-economic groupings of households, and are based on cross-sectional data (rather than time series data). He feels that to a large extent the meager use of and loss of enthusiasm for some of these models is due to the fact that the formulation of the urban development problem has changed—away from planning metropolitan areas or cohesive functional units and toward solving problems people have with housing, employment, discrimination, etc. Hemmens thinks the framework within which future models are developed and used will be more important than better data and more sophisticated techniques.

National Bureau of Standards Survey

Some urban model surveys have been performed for special types of users and have employed specific and difficult-to-satisfy selection criteria. An example of this type of survey is the one prepared for the HUD Model Cities program by the Institute for Creative Studies and the Technical Analysis Division of the National Bureau of Standards. Using subjective "ideal model" criteria of: comprehensiveness, high disaggregation, short-run time frame, analytic (i.e., behavioral), predictive, easy to use, and operational, it is little wonder that the study found no acceptable urban model. Yet the study recommended the use of a model (despite known deficiencies) for Model Cities evaluation purposes because of (1) the assistance it could provide in the evaluation process, (2) the organization the model would bring to data collection activities, and (3) the educational role the model could play with local citizens and officials.

For Model Cities purposes the study concluded that the conventional urban spatial models and urban input-output models are less fruitful starting points than the approaches suggested by urban dynamics models, urban gaming models, and aggregated impact models.

Nagelberg and Little Survey

This survey recognized three basic types of urban models: computer simulations, games, and gaming simulations. For the nineteen models covered in this survey, the authors attempted to contrast the models with respect to purpose, uses, costs of development, major assumptions, numbers and types of parameters, computer requirements, etc. For many of the models, inadequate information makes it difficult to actually make cross-model comparisons. An interesting effort was made to provide a geneology of the urban simulation models, but inaccuracies made the effort less than a full success.

Sweet Book

This edited work contains an interesting survey article by Leslie King, the 1960 residential model description by John Herbert and Benjamin Stevens, the SEWRPC model description by Kenneth Schlager, a description of a land use allocation model applied to Franklin County, Ohio by William Habig. Other chapters of the book are devoted to new concepts in urban-structure models.

An appendix of this book shows the results of a survey on computer utilization and model development by metropolitan planning agencies. A 40% response from 226 agencies surveyed yielded the finding that only 51 of the 91 responding agencies were using or planned to use computers. Only 26 of these 51 agencies were currently using urban development models (17 for transportation and 9 for land use). Only 8 of these agencies originally developed the model they were using. Most of the respondents did advocate the development of additional urban models that would (1) provide better forecasts, (2) be cheaper to run, and (3) quantify the impacts of proposed land use changes. They favored urban development models that could use existing data files.

Goldstein and Moses Review Article

This recent economics-oriented survey of urban growth, land use, and market process models is noteworthy for its fine treatment of the theoretical models of Mills (1972) and Muth (1969), its comparison of the comparative statistics approach (Lowry, 1965) and the dynamic approach (NBER, 1969) and its very comprehensive (438 items) list of references.

Goldstein and Moses review various theoretical models of urban processes at greater length than is possible in this chapter. The reader is referred to their work.

To give a rough idea of the magnitude of previous efforts with regard to urban modeling, the Regional Environmental Systems Analysis Program at the Oak Ridge National Laboratory has collected *5000* abstracts of literature on the development and use of "mathematic models capable of simulating the economic, demographic, societal, ecological, and land-use responses of a geographical regional to alternative policy decisions" (Myers, 1973).

European Surveys

Michael Batty (1970) from the Geography Department at the University of Reading has provided one of the best reviews of land use modeling in Britain. Most of the European models have been derivatives of Lowry's Model of Metropolis (1964). A. G. Wilson (1969) of the Center for Environmental Studies has added residential complexity to the Lowry scheme. Batty (1970b) himself has applied a modified Lowry model to the Nottinghamshire-Derbyshire subregion and to the Central Lancashire subregion (Batty, 1970c). These two plus three other applications of the modified Lowry are dealt with in some detail by Batty in this excellent survey.

William Goldner (1971) at the University of California at Berkeley reviewed the foreign applications of the Lowry model derivatives and contrasted these to the meager applications of the Lowry model in the U.S. Goldner discusses the application by Cripps and Foote (1969) in Bedfore County, Echenique et al. (1969) in Reading and Stevenage, Music and Barber (1970) in Ljubljana, Yugoslavia, and the two applications by Batty mentioned above.

What is clear from both Batty and Goldner is that only the Lowry Model derivatives, and to a lesser degree the Lakshmanan-Hansen model, of all the U.S. developed land use models, have found much use abroad, and most of this use has been in Britain.

IV. SELECTED URBAN DEVELOPMENT MODELS

From the preceding review, it is apparent that urban development models were a growth industry particularly in the 1960's. Quite a number of models that vary in theme, policy scope, theoretical structure and empirical content emerged in this era.

These urban development models essentially allocate elements of urban growth (firms, economic activities, people) to geographical subunits of a metropolis. However, they differ in the criteria used for such allocations. Some models are guided by past patterns of development in the urban region; some utilize microeconomic behavioral notions of location; others formulate optimization rules; still other models seem to rely on what may be described as intuitively attractive ad hoc propositions.

Some of the models describe only one major sector of urban growth; for example, the residential sector (Hansen 1959, Herbert and Stevens 1960, Donnelly, Chapin 1965, A. D. Little 1966, NBER 1972) or the retail sector (Lakshmanan and Hansen 1965, Berry 1965) or the industrial sector (Putman 1967, Goldberg 1967, Seidman 1969, Rose 1967). Such models dealing with one sector can be described as *partial* models (Batty 1970a) that may form components or submodels of general or comprehensive models.

General or comprehensive models deal with the broader range of urban activities than one sector (Lowry 1964, Hill 1965, Seidman 1964, Lakshmanan 1968, Forrester 1969).

Further, the models evidence a variety of theoretical approaches. Some formulate other models using microeconomic notions of decision making in the urban land making (Herbert Stevens, NBER, Chapin). Some build upon probabilistic notions of interaction (Lakshmanan and Hansen) or intervening opportunity models (Lathrop and Hamburg, 1965). Some use gaming-simulation approaches (Feldt 1972, U.S. EPA 1972a and b).

Others use optimizing models of location (Schlager 1965).

In this chapter we present a selective illustration of the dominant themes and trends in urban development modeling. We begin by recognizing at least two broad groups of models.

One group of models representing the highest level of abstraction is the set of mathematical models that describe urban development in analytical terms. These analytical models use a variety of mathematical models—sets of simultaneous equations, linear programming formulations, differential equations and so on. Most urban models are of this analytical variety. We shall further classify them into subgroups—partial and general or comprehensive models.

Another broad class of models is simulation-gaming. A simulation-gaming model essentially is a game defining the nature of interactions between players assuming various decision roles and a simulation model that analytically translates the consequences of the gaming choices. Generally speaking, the players set out certain decision roles guided by a set of rules, and their output is translated in the simulation phase into conditions that describe the outcomes at the end of play of the game. These conditions set the stage for the next play. A number of these gaming-simulation models have been built largely in response to a perceived need for training and teaching. The educational objective of this class of models is, in fact, their dominant objective.

From the above discussion we can classify urban models as follows:

1. Analytical Models—partial
   These are mathematical models that deal with one sector of urban development such as residential or commercial.
2. Analytical Models—general or comprehensive
   These are mathematical models that describe two or more sectors of urban activities.
3. Simulation-Gaming Models.

Using this classification, Table 1 lists the models that are reviewed next.

Table 1. Models by Class Discussed in This Chapter

Class of Model and Name	Brief Description
A. *Analytical Models*	
Partial	
Residential Sector (NBER Model)	A dynamic economic model of the housing market.
Retail Sector (The Retail Market Potential Model)	An operational model of retail shopping activity.
Comprehensive	
Lowry Model and its derivatives (TOMM, PLUM, Batty)	A model that describes the location of residences and nonbasic employment—several prototypes of the Lowry Model development.
Urban Dynamics Model	A model of long run growth processes.
B. *Simulation-Gaming Models*	
Community Land Use Game (CLUG)	Relatively simple table game—closely tied to economic location theory.
RIVER BASIN MODEL	Players represent economic, social, and governmental decision makers. Uses the computer extensively to simulate migration, housing, employment, transportation, shopping and time allocation activities.
APEX	Players represent major political and industrial decision makers in an actual urban area (Lansing, Michigan). Computer simulations of new industrial growth, housing occupancy, voting, and pollution impacts.

Partial Analytic Models

NBER Model—The Residential Sector

Models of residential development have attracted greater research than any other type. From the early work of Walter Hansen in Washington, D. C. and that of Herbert and Stevens for Philadelphia to the current efforts. A number of residential development models have been developed either as partial or submodels of comprehensive models. One of the more ambitious of these models that incorporates some features of the Herbert and Stevens model is an ongoing effort at the National Bureau of Economic Research (Ingram, Kain and Ginn, 1972). The NBER model is an ongoing effort developing various prototypes such as DETROIT, PITTSBURGH I and II. Except where explicitly noted, however, the model described here will be the one in the Ingram, et al. book (1972).

The National Bureau of Economic Research model—more than any of the other models surveyed in this chapter—focuses specifically on the housing market and describes that market in considerable detail. The location of employment in both local service and export industries is taken as exogenous, and the principal variables determined within the model are the spatial distribution of households throughout the urban area and the types and prices of housing which they consume. The model is a short run dynamic simulation model which starts with a given allocation of households and housing units at a starting period, takes demographic and industrial changes as exogenous, and models the resulting changes in the location and housing of families over time. The model carefully distinguishes between the demand side and the supply side of the housing market and portrays the way in which both sides respond to prices (see Figure 2). Hence, it will be convenient to describe the model by beginning with the demand side, then discussing the supply side, and then describing the process by which prices are set.

In any given period, the model takes as given the allocation of households to housing which was made in the previous period. The active participants on the demand side of the model in the new period, then, consist of those families which move during the period either as

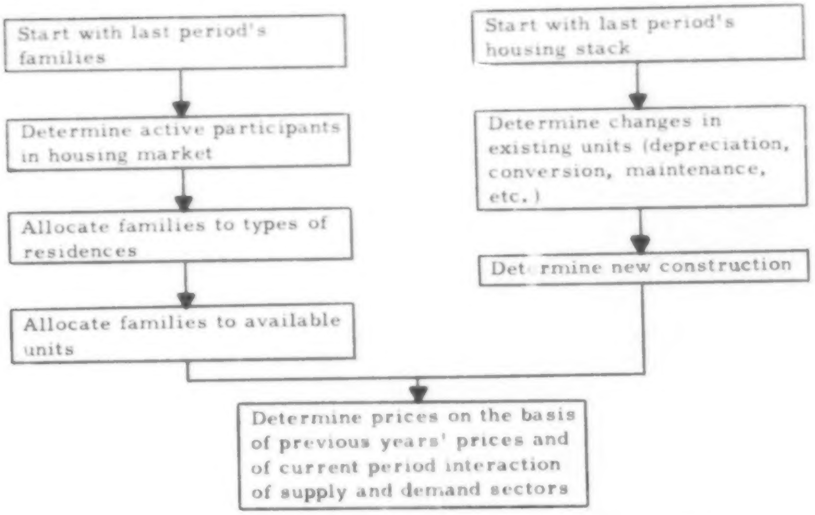

Start with last period's families
Determine active participants in housing market
Allocate families to types of residences
Allocate families to available units
Start with last period's housing stack
Determine changes in existing units (depreciation, conversion, maintenance, etc.)
Determine new construction
Determine prices on the basis of previous years' prices and of current period interaction of supply and demand sectors

FIGURE 2—*National Bureau of Economic Research Model Processes*

a result of new household formation, migration into the city, changes in stage in life cycle or changes in employment. These families are divided into different family types on the basis of income, family size, age of head, and place of employment of head. The model must allocate the families among 27 different possible types of dwelling units and also among the 44 different geographical residence zones in the city. For instance, one class of families consists of those who have less than $10,000 of income, which have 3 or 4 persons, and which have a family head less than 30 years old who works in a specific area. The model must allocate families in this class to housing types (e.g. single family-large lot-high quality or duplex-row, or apartment-small structure-low quality, etc.) in different residential areas. This allocation is done as a two-step process. First the families are allocated to one of the 27 housing types on the basis of demand equations which are a function of the prices of the different kinds of housing. Then, having allocated the families to specific housing types, the model allocates the families to available housing units of these types in different parts of the city using a linear programming algorithm which minimizes the sum of families' transport costs. Thus the demand sector of the model takes housing supply and prices as given and allocates families on the basis of prices and of family commuting costs.

The supply side of the model is based on the assumption of profit maximizing behavior in the housing industry. There are a number of basic decisions which housing suppliers must make. They must decide whether to build new buildings, whether to convert buildings from one use to another, and how much maintenance expenditure to devote to existing buildings. These decisions are made in the model on the basis of expected profitability. Essentially, the supply side of the model takes as given the prices for different kinds of housing units and the costs of alternative investment decisions—i.e., new buildings, conversion of buildings, etc.—and then selects investment projects which will be profitable. This selection is done subject to a number of constraints including both zoning laws and also limits in the amount of inputs available for housing construction in any given time period.

Thus both the demand side and the supply side of the model use prices as inputs in allocating families to housing and in generating housing investment decisions. The final step in describing the structure of the model is therefore to discuss how the prices themselves are determined. Essentially, prices are calculated using the "duals" from the linear program which was used to allocate families to different residential zones. This is a somewhat sophisticated technique and requires careful elaboration. As was described above, the demand sector of the model after it assigns families to specific housing types, uses linear programming techniques to allocate the families to different residential zones in such a way as to minimize commuting costs. It is a feature of linear programming techniques that they generate a set of values, often called "duals," which—in the present case—can be interpreted as indicating the travel cost savings which would have resulted from having one additional dwelling unit of each type in each residential zone. For instance, one of the dual variables can be used to estimate the commuting cost savings which would have resulted in having—say—one extra duplex-row type dwelling unit in a specific area. This, then, can be interpreted as the amount of extra money which a family would have been willing to pay for such a unit in that location in addition to what it actually paid for the dwelling unit that it was actually assigned to. And these values, in turn, can be converted into a set of relative single-period housing prices for different housing types in different locations. The prices actually used as inputs in the supply and demand parts of the model are expected prices based on weighted averages of single-period prices for previous periods.

This completes our basic description of the structure of the model. The model is basically a short run, dynamic model. It computes the value of its variables in a given period not on the basis of a set of long term equilibrium conditions but rather on the basis of the variables' values in the previous period and of simulated short run decision making by demanders and suppliers in the market.

While the creators of the model are careful to point out that they have not modeled a particular city, e.g. Detroit, in any literal sense, the parameters of many of the equations for the model—particularly on the demand side—are estimated using econometric analysis of data from the Detroit region. Detroit data is also used to generate initial values for the variables in the model.

Appraisal

Rather than creating a tool for predicting detailed land use patterns in specific geographical areas of the city, the authors of the NBER model are more concerned with creating a model which will shed light on the underlying dynamic forces shaping the structure of the city and which will also be useful in analyzing the effects of alternative government policies, particularly those affecting housing markets. They emphasize in their report, however, that their research is far from completed and that the present model is to be taken more as a progress report rather than as the presentation of a final version of work which achieves these objectives. Nevertheless, it may be of interest to discuss their progress so far, as represented by the current model, in light of the evaluatory criteria outlined earlier in this chapter.

As we have noted before, the model represents an explicit attempt to incorporate more economic theory than that embodied in the typical empirical model while at the same time retaining more geographical disaggregation and more careful attention to calibration with data than the theoretical models such as those of Muth or Mills. Hence, we shall examine the model with regard both to its theoretical and empirical content.

It is clear that the model does in fact go far beyond most of the empirical land use models with regard to the degree to which it embodies theoretical relationships. In particular, the explicit modeling of market behavior and prices on both the demand and supply sides of the market is an interesting contribution. It should be pointed out, however, that the necessity of keeping the model computationally feasible has forced a number of compromises with theoretical consistency. For instance, as we have described, the model represents families' housing choice behavior as two distinct, sequential steps: first the family chooses a type of dwelling unit; then it chooses a specific residential area. In theory and fact, of course, these two decisions are made simultaneously. Another example of compromise with pure theory occurs on the supply side of the model where the algorithm which the model uses to select profitable housing investments is only an approximation to the theoretically-expected profit maximization behavior.

With regard to calibration with data, the creators of the model have had mixed success, also. Data limitations have prevented serious attempts to statistically estimate the parameters of many of the equations on the supply side of the model. Substantially more effort has been devoted to econometrically estimating parameters of equations on the demand side. But that there are difficulties in this work is indicated by the fact that many of the equations which they estimate do not show a statistically significant relationship between the amount of a given type of housing demanded and its price.

It should also be pointed out that the current version of the model leaves out a number of key relationships which any complete model of even residential location must take into account. For instance, as we have seen, all employment is taken as completely exogenous. Also the effects of race and externalities between different types of residential land use are not included in the model.

In summary, then, the model clearly represents an interesting attempt to combine some of the best features of previous models, but—as its authors themselves are careful to point out—in its current version it is clearly not sufficiently well-developed for use with any degree of confidence in policy analysis.

The Market Potential Model—The Retail Sector

The retail market potential model was developed to describe the sales of each of a set of large shopping centers in Baltimore. This model is descended from an earlier one (Lakshmanan, 1965) and builds on the probabilistic notions of shopping spatial interaction (Huff, 1962).

The model states that the size and the number of retail establishments in an area is a function of consumer shopping expenditures accessible to that area. The retail center in zone j attracts consumer dollars (from consumers in zone c):

- in direct proportion to the consumer expenditures
- in direct proportion to its size
- in inverse proportion to distance to consumers, and
- in inverse proportion to competition

The total sales at the shopping center in zone j is the sum of all consumer dollars attracted in this manner from all zones where consumers live.

A simple econometric model relates consumer expenditures for shopping goods as a function of family income. The use of this model and data on zonal population and family income generated the consumer shopping expenditures by zone. The size of the shopping center was measured by floor space of the shopping center.

Appraisal

The market potential model was utilized to generate annual sales of shopping centers in the Baltimore Region. Sales estimated by the model of a few large shopping centers were within 5% of the actual sales at these centers as obtained from the sales tax records. A more general assessment of the model was carried out by using it to estimate shopping trips from residential zones to shopping centers and verifying these against origin-destination survey data. The model's ability to predict shopping interactions in this fashion was noteworthy.

The model was used to locate a set of shopping centers in the future that balanced their profitability against the need to maximize consumer access to them.

The data required to develop and apply this model are generally available in most transportation studies. The latter generate data on population and family income by subarea. The factor that describes the spatial attenuation of shopping interaction from a center can be derived from the shopping trip distribution data via a gravity model. In this sense, this model can be implemented with the data available in transportation studies.

This model utilized probabilistic notions of consumer spatial behavior rather than explicit notions of market processes. It is possible to view the gravity type model as a shortcut way of capturing the outcome of the explicit market processes.

Another aspect of this model is the rather simplistic representation of the attractive power of a shopping center. Size is only one measure; the composition and order of shopping goods and their "image" to consumers are other dimensions of the attractiveness. Size of the shopping center may be, to some degree, correlated with these dimensions, but there is definite ground for refinement here.

There are quite a few versions of this model that have been developed in Britain and elsewhere. McLoughlin, Nix and Foot (1966) used this model to estimate the impact of a large shopping center on other shopping centers in northwest England. These authors modified the variable that represents the attractiveness of the shopping center (size or floor space in the Baltimore Model). They developed and used a functional index complied from data measuring the variety of retail stores in each center.

Two other variations of the model are documented for Oxford (Black, 1966) and Lewisham (Rhodes and Whitaker, 1967). The major modifications in these applications were to the attractiveness variable.

Comprehensive Analytical Models

The Lowry Model

In this section we shall present in detail the model by Ira Lowry (1964). Lowry's model is appropriate for our discussion here because it illustrates a number of important strengths and weaknesses which characterize most of the empirical land use models discussed in the surveys reviewed in Section III of this chapter. Although only Lowry's model will be discussed in detail in this section, various derivatives of this general type such as the Projective Land Use Model (PLUM) of the San Francisco region, will be mentioned briefly in order to illustrate some recent developments of this general area of urban model building.

Lowry's model is calibrated to data collected in the Pittsburgh metropolitan area. He divides the Pittsburgh metropolis into separate geographical districts each with an area of approximately one square mile. The model distinguishes between two kinds of employers: export industries which produce goods and services sold outside of the metropolitan area, and local service industries which produce goods and services consumed locally. The export sector is further disaggregated into several major industries, while the local service sector is disaggregated into three different kinds of local service clusters depending on the market area served by each kind of cluster (e.g. one category—"neighborhood facilities"—needs only a relatively small number of customers to operate efficiently, while the other categories—"local facilities" and metropolitan facilities"—need progressively greater market sizes per cluster.

The model takes the location and amounts of employment in the export sector industries as exogenous and seeks to allocate local service and residential activity to each of the small districts in the urban area. The procedure for doing this is an iterative one and proceeds as follows (see also Figure 3): As indicated above, the model starts with the exogenously-determined export industry employment levels in each district. The next step is to allocate export industry workers from each of the districts to residences. Essentially this is done using what is referred to in the transportation chapter of this book as a "gravity model"—workers who are employed in one region are allocated to residences in other regions on the basis of the accessibility (in Lowry's model, principally just distance) between the two districts. The exact parameters of the allocation formula are determined using Pittsburgh data. For instance, if the data showed that in the Pittsburgh region approximately 15% of managerial workers lived in districts which were two miles distant from their place of work, then the model would allocate 15% of such workers in any given district to residences in districts two miles away.

So far, then, we have seen how the model makes an initial allocation of export workers' residences. The next step is to bring in the local service sector. This step begins by estimating the market potential for each of the types of local service industry in each of the districts. This too is done with what is essentially a gravity model. The number of residents of district "i" who will shop in district "j" is assumed to be an inverse function of the distance between the two districts where, again, the parameters of the function are determined by the available data. After this is done, for each j, the total number of shoppers coming to district j to shop is computed by adding up the number

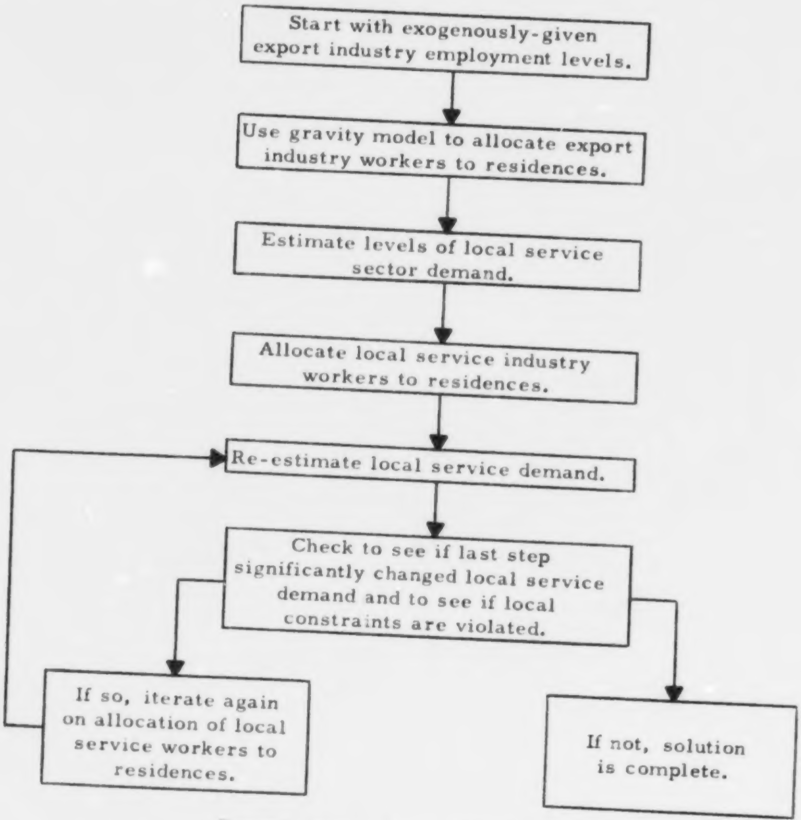

Start with exogenously-given export industry employment levels.
Use gravity model to allocate export industry workers to residences.
Estimate levels of local service sector demand.
Allocate local service industry workers to residences.
Re-estimate local service demand.
Check to see if last step significantly changed local service demand and to see if local constraints are violated.
If so, iterate again on allocation of local service workers to residences.
If not, solution is complete.

FIGURE 3—*Steps in Solving Lowry Model*

coming from each district (including j itself). Finally, in order to get total market potential in district j, the number of shoppers just computed is added to an estimate of the number of shopping trips generated by people who work in district j on the theory that workers may make short trips from their work place to purchase local services. The market potential generated by workers of a given type is simply taken as proportional to the total numbers of such workers in the region.

After market potential in each district is estimated, local service employment is allocated to each district in proportion to the market potential, except that for each type of local service industry a minimum size constraint is imposed. If the market potential in a given district is not at least as great as the minimum size constraint, no local service employment is created in that district, and shoppers are sent to nearby districts. (This minimum size constraint is necessary, of course, to simulate economies of scale in some way—without it there would be some fraction of a major hospital, for instance, in every small district.)

The allocation of local service employment, then, adds more workers to each district, and these are allocated to residences in the various neighboring districts on the basis of the same gravity model used to allocate export sector workers to residences. The addition of more workers and more residences to the model however, changes the previously computed local service industry market potential and hence requires more local services in each district. Therefore the model proceeds by iterating again on the previous procedure. Specifically, new local service market potentials are determined on the basis of the initial allocation of local service employment, and these market potentials are used to create a new allocation of local service employment, and a new allocation of residences for local service workers. This leads to yet a third estimate of market potentials, and the model thus proceeds iteratively until it converges on a stable solution where the computed values of market potentials do not change significantly between iterations.

Additional constraints are also put on the model to ensure that the total amount of activity allocated to a district does not exceed the capacity of useable land in the district. Specifically, the model uses data from the Pittsburgh area to determine the number of square feet of land needed per worker employed in each of the types of industry. Then land is allocated to each industry in each district on the basis of this data and the remaining land is allocated to residences as a residual. However, two constraints are imposed in this process. One is that total land used for export and local service industries cannot exceed total land available in a district, and the second is that the density of residences per square foot of land in a district must be below a maximum value set by the model on the basis of existing residential densities in the Pittsburgh region. If either of these constraints is not met in a district, some of its residential or local service activity is allocated elsewhere.

As indicated earlier, there are quite a few prototypes of the Lowry model both in the U. S. and in the U. K. These prototypes run the gamut from conceptual modifications (Garin-Rigers, 1966) through experimental-conceptual (Crecine, 1964; CREVE, 1965; Echinique, 1969) to operational-experimental-conceptual versions (PLUM, Batty and Ljubijana). Goldner (1971) has provided an excellent survey of these. We shall describe here only the PLUM model which appears to be applied to a number of metropolitan areas.

The PLUM model follows Lowry's basic approach in dividing up the urban area into zones and estimating employment and transportation patterns in and between zones, but it adds a number of refinements. For instance, in computing transportation times, more careful attention is given in the PLUM model to estimating the impact of differences in availability of transport systems between different zones. The fact that trip times may vary at different times of the day with different amounts of congestion is also brought into the model. Furthermore, the PLUM model devotes greater attention to modeling acreage requirements for residential land rather than simply taking residential densities as a residual as Lowry does. Too, it allows for the possibility that in different parts of the city there may be different numbers of workers per household and also different numbers of people per household. Somewhat more sophisticated functional forms are used to replace Lowry's gravity model specifications in allocating workers to zones.

Appraisal

We have described the Lowry empirical land use model in considerable detail and we have shown how more recent work in the area, as illustrated by the PLUM model, builds upon the Lowry approach while making substantial refinements on how the approach is actually carried out. We shall now discuss the strengths and weaknesses of the Lowry model—and to some extent of the general group of empirical land use models—in light of the evaluatory criteria developed in an earlier section. As was pointed out, however, urban models can best be evaluated in comparison with the objectives which they seek to accomplish, and it may therefore be appropriate to consider explicitly the objectives of this type of model. Of the four objectives identified in our previous discussion, the two which these models are most concerned with are prediction and policy analysis. The designers of these models have been specifically interested in predicting land use patterns, and the way in which models like Lowry's can be applied in this way is reasonably straightforward. For instance, in the case of the Lowry model, once an independent forecast of the location of export industry employment at some point in time is made, it is then possible to use the model to predict other types of land use. It is also reasonably straightforward to see how models of this type can be used for policy analysis. For instance—though Lowry does not use it in this way in the cited report—the model could be used to investigate the possible effects of an urban renewal project by constraining the districts in the urban renewal area to have the type of new land use proposed for the project. It is also possible to use some models of this type—though not Lowry's—to analyze partial effects of alternative transportation policies. This can be done in models where the gravity relationship which is used to allocate residence places and work trips is made to depend on both distance and transportation costs. Given this specification, the effects of changed transport systems can be modeled as changes in transport costs.

The two basic objectives of this kind of model, then, are prediction and policy analysis. Two other objectives—optimization and understanding basic structural relationships shaping the city are less important in these models.

Consistency With Data

Turning now to an appraisal of these models, one of our proposed criteria is consistency with data, and on this criterion models of this type appear to be quite impressive. Given the planning objectives of these models, it has been necessary that they be fitted as closely as possible to actual data, and the creators of these planning study models have made careful efforts to use real world data to a far greater extent than the builders of most of the other models which will be surveyed below. In some cases literally millions of dollars have been spent in gathering data and processing it in connection with the models. It is worth pointing out, however, that in spite of these efforts many data problems remain. One of the most important of these difficulties is what has been referred to as "cross-sectional bias" (Brown, et al., 1972). Because of data limitations mentioned in our earlier discussion of urban modeling, many of these planning study models have found it necessary to generate their own data through surveys and sampling techniques. This creates difficulties, however, because land use patterns at one point in time may not adequately reflect current market forces due to extremely long adjustment lags in urban systems. Densities of residential patterns, for instance may be strongly influenced by buildings created many decades previously, and hence, existing land use patterns may not be entirely accurate bases from which to extrapolate future patterns.

Another problem concerning data which empirical land use models face is that, even with their major data collection efforts, the accuracy of the calibration of these models with data is often substantially hindered by the general problems in working with urban data which were discussed earlier. This creates an evaluation problem in that it has proved to be extremely difficult to develop criteria with which to judge the accuracy of the models. At the very minimum, of course, they should be expected to generate results which have at least some resemblance to the city being modeled. And indeed they do. But this minimum resemblance is almost guaranteed by the structure of the models—in the Lowry model, for instance, by determining export industry employment exogenously,—and, beyond the minimum resemblance criterion, it is difficult to develop reasonable and useful indices of the degree to which the models fit the real world.

In spite of these data problems, however, it is clear that the empirical land use models go further toward meeting the consistency with data criterion than do most of the other models to be surveyed. However, when judged against one of the other criteria identified previously—consistency with theory—the empirical land use models do not compare as favorably with other types of urban models.

Consistency With Theory

One theoretical limitation of most of the models of this type is that they do not explicitly portray market processes in urban land markets with any degree of thoroughness. Land in different parts of the metropolitan region has different values depending on the relative desirability of the alternative parts of the region for various kinds of economic activity. Presumably, potential land users are influenced by these prices in making their location decisions, and indeed standard economic theory would suggest that these prices play a key role in the land allocation process. In many of the empirical land use models, however, prices do not appear at all—as for instance in the Lowry model—or, if they do appear, they appear in a largely superficial way. It should be remembered, of course, that in terms of the objectives of the creators of models of this type, suppression of market processes may be a reasonable procedure, since loss of detail in this area may make it feasible to have more detail in other respects—for instance in the geographical disaggregation which is these models' principal objectives. Indeed, for some purposes, gravity type models like the one Lowry uses can be looked upon as a short-cut way of representing the final outcome of the market process, and Lowry discusses this interpretation in his report. Nevertheless, failure to carefully model market processes certainly weakens the theoretical underpinnings of these models.

Another criterion with which to judge models is that of whether given the purposes of the models, they include the variables and relationships which are intrinsically necessary in accomplishing their objectives. The basic objective of these models has been their use for disaggregated local planning, and the models have indeed been successful in achieving a good deal of geographical disaggregation. There is, however, at least one respect in which they have omitted a relationship which may be intrinsically necessary for their planning objectives. As noted above, many of these models lack a careful attempt to describe the locational decisions of export industries. Now, for the purposes of planning on a relatively short-run planning horizon—say for the next 5 years—this may be an entirely appropriate simplifying technique. However, one objective of these models has been their possible use as an aid in transportation investment planning. And this creates some difficulty because transportation investments—e.g. roads or train tracks—are often extremely long-lived. In considering the effects of such investments it may not be appropriate to take the location of export sector activity as exogenous. This is especially true since access to transportation may well be a key factor in plant location decisions. Therefore, empirical land use models have often been criticized for the fact that, even though some of them have been intended to be used as an aid in transportation planning, they have concentrated mostly on the effects of industrial location on transportation needs while largely ignoring the feedback effects running from the transportation system to the location of industrial activity.

Another characteristic of many of these models which it may be appropriate to mention here is that for the most part they have ignored the existence of different racial groups and of racial segregation.

In summary, then, it can be said that the empirical land use models can best be understood if considered in light of their principle objective—i.e. detailed land use planning. The suppression of certain variables and relationships in these models and their limited use of theory have apparently been necessary in order to make it feasible to attain the geographical disaggregation and attention to data which their planning objective required. Nevertheless, even taken on their own terms, a number of criticisms of models of this type can be made. In particular, their failure to adequately consider the effects of planning decisions on export industry location limits their usefulness. So too does their failure to deal with racial discrimination.

Since many of these models were originally undertaken with the goal of making them operational as planning devices, perhaps the most reasonable ultimate criterion on which to judge them is whether or not they have proved useful for this purpose. Here, too, however, the record is mixed. In some cases, despite huge expenditures of time and financial resources, the models have not been made fully operational as planning devices and have had to be abandoned by specific planning projects. An example of this, for instance, is the Penn-Jersey study of the Philadelphia region (Levin and Abend, 1971). In other cases, however, some use has been made of the models in planning applications. For instance, the PLUM model described above has been made operational, and so, too, have a number of similar models designed for planning applications in England. (For a description of these English applications, see Goldner, 1971.) Of course, even in these cases where models of this type have been made operational, there is no way to judge with any certainty whether the final planning decisions have been improved by the use of the model. It is clear that no current model of this type can be used mechanically by itself for planning purposes—too many variables have, of necessity, been omitted. Some planners have found them quite useful, however, in suggesting general planning patterns and in checking for overall consistency of plans. Other users, on the other hand, have felt that they have added relatively little beyond that which could have been accomplished by good planning judgment. In any event, the overall record of these models is a mixed one and it seems safe to say that not all of the hopes for such models which were held in the early 1960's have in fact been fulfilled.

The Urban Dynamics Model

The next model which we will consider is the Urban Dynamics Model developed by Forrester (1969). This model attempts to explore questions of long run urban growth and stagnation.

Forrester begins his model by identifying three broad classes of potential workers (managers, laborers, and the underemployed), three broad classes of housing (premium housing, worker housing, and housing for the underemployed), and three general types of businesses (growing industries, mature industries, and declining industries). A public sector with taxes is also built into the model. Changes in these variables—as well as some additional, auxiliary variables—are then modeled with an interactive, non-linear, dynamic equation system which is meant to represent the economy of a typical city. For instance, the rate of migration of underemployed persons into the city depends in the model on an index of the attractiveness of the city to such people which, in turn, depends on the availability of housing and of new jobs. Or, to take another example, the amount of worker housing in the system depends on the amount of premium housing becoming obsolete and on the rate of new construction of worker housing. And this latter variable, in turn, is a function of a number of other variables including the size of the worker population, the availability of land, and tax considerations.

Altogether, the model consists of some 154 equations which model relationships of this sort between the variables. Forrester's model is a dynamic model which starts with a set of initial conditions and simulates the behavior of the system over time. In terms of the span covered, Forrester is considerably more ambitious than other urban models—he attempts to model changes in cities over several hundred years.

On the basis of the relationship established in his equations, as well as a set of assumed parameters and initial conditions, Forrester presents the results of three sets of simulation runs of the model. The first set of simulations deals with the life cycle of cities. Within a 250 year time frame, the simulated city grows slowly for the first 50 years, then rapidly for the next several decades until about the 100th year when population, housing, and business are at a peak and all available land has been consumed. From this point on the forces that generated the growth lead to rapid decline until relative equilibrium and stagnation are reached in year 175. This equilibrium remains stable throughout the remainder of the run. Forrester takes these results to be relatively representative of historical city growth processes. He further asserts that this representativeness is a measure of the validity of the assumptions and structure of the model.

The next series of studies which Forrester presents is an analysis of the effects of various programs "to save the cities" as they would operate given the assumptions of the model. Starting from the equilibrium-stagnation point developed in the first run, Forrester alters the assumptions and parameters in the model to represent in turn a job program, a training program, a financial aid program, and a low-income housing construction program. The job program has very little impact on the system. It results in slightly higher numbers of low-income population and low-income housing but lower numbers of working-class housing. Higher taxes and decreased new construction also results. The training program has similarly little impact on the city. Its most dramatic effect is that it transforms the city into a conduit for training low-income populations. People from outside the city are trained in the city and then return to the outside environment when they find that job opportunities are limited by the rapid increase in people with working class skills. The fiscal aid program results in the city being more attractive to low-income people, less attractive to new housing and business and consequently requiring a higher tax rate. This counterintuitive result is caused by the increasing demands placed on the city as a result of the shifts in population, housing, and business. The low income housing program has rather major effects which are detrimental to the city. The program leads to an influx of underemployed workers and rapid departure of working and upper class populations as well as new construction. Forrester concludes that this series of policy simulation runs is indicative of the counterintuitive nature of the urban system.

As a final step, Forrester experiments with other possible policy options for improving the condition of the city, including worker-housing construction, premium-housing construction, new-enterprise construction, declining-industry demolition, slum-housing demolition and discouraging housing construction (i.e., slum-housing demolition and depression of worker-housing construction). Of these seven alternatives the results of running the model are most unsuccessful for the city with the housing construction programs and most successful for the combined slum demolition-new enterprise construction program. To Forrester these results confirm his theory that many city problems are the result of an economic imbalance between housing and business which cannot possibly be redressed through housing programs.

As a result of these experiments Forrester has developed a series of conclusions as to the behavioral characteristics of complex systems as modeled in Urban Dynamics. First, as suggested earlier, he concludes that they are counterintuitive. That is, they often respond to policies in a way opposite to that suggested by intuition or observation. Second, after performing additional tests on the model, he concludes that the results are relatively insensitive to changes in the parameters of the model. This suggests that the structure of the model is of considerably more importance than its assumed parameters or calibration. Third, he argues that the experimental runs described above suggest that the system is relatively unresponsive to policy changes. Fourth, Forrester notes that the system as modeled is able to counteract and compensate for externally applied policy measures. Finally, he notes that the results of the model indicate a tendency toward low performance (eventual stagnation) under a wide variety of circumstances. While these five conclusions do not provide a basis for optimism they do, according to Forrester, suggest an important new way for policy makers to look at a system which they are attempting to influence and control.

As was the case with previous models discussed in this survey, it is useful to consider the goals which Forrester's work seeks to achieve before attempting an evaluation of the model itself. Of the four general objectives discussed at the beginning of this chapter the Urban Dynamics model seems most concerned with improving our understanding of the basic forces shaping urban areas. Some qualifications concerning this must be made, however, on the basis of the general nature of the model. Many of the relationships of the model are not solely economic in nature. For instance, Forrester attempts to include variables concerning perception and attractiveness which are determined by both economic and noneconomic forces. In attempting to include relationships of this type, Forrester has ignored questions of urban form or spatial location in order to keep the model manageable.

The model clearly has a general and "macro" orientation, and the relevant policy issues with which this model deals are more questions of direction than detail. Except in the most general sense this model does not seek to be predictive, nor is optimization one of its aims.

Appraisal

Given its goals, we may evaluate the model in terms of the criteria we have specified. First we shall consider the extent to which crucial variables have been included or excluded. As noted earlier, the variables included may be generally classified as those concerning population, housing, and business. Within these general categories Forrester attempts to specify variables concerned with the political, physical, social, psychological, and economic nature of the city. Forrester's model is a large one with a great number of variables. However, its lack of specific attention to spatial location limits its usefulness for analyzing some important urban phenomena such as suburbanization. The model also ignores the influence of transportation technology and technological change of all kinds. With regard to our simplicity criterion, the many equations in Forrester's model, while making its operations somewhat intriguing, do not lend themselves to ready analysis. Those relationships or effects which are really important to the model's underlying theory are hidden in the complex form in which the model is presented.

Our third criterion is consistency with theory. Forrester makes no attempt in the Urban Dynamics model to use the existing theory of urban development. He notes in his preface that in specifying the relationships of his model he consulted persons with practical experience in urban affairs and drew principally on their perceptions and experiences. The book contains virtually no references to existing theoretical literature in economics, political science, urban planning or other professional disciplines.

Our final criterion involves the model's consistency with data. Forrester does not seek to use data in the specification of his equations. He uses data analysis neither in fitting equations nor in testing generally for the existence of relationships. All the parameters in the model are made up on the basis of what seemed intuitively plausible and of what generated apparently plausible simulation results.

In summary, this discussion of the Urban Dynamics model in terms of our evaluation criteria leads to the conclusion that while the model presents an interesting view concerning the processes involved in city growth, its validity is difficult to verify because it is not carefully related to the current state of empirical and theoretical knowledge.

The System Dynamics group at the Massachusetts Institute of Technology is applying the Urban Dynamics Model to Lowell, Massachusetts with HUD support (HUD Contract H2000-R). Early indications are that Lowell may be one of the places in the U. S. that might conform to the pattern of growth and decline forecasted by the Urban Dynamics Model.

Simulation-Gaming Models

We turn now to a class of models which have generated a good deal of interest: urban simulation-gaming models. Various models of this type (Envirometrics, 1971) have been developed with different degrees of emphasis placed on the gaming and simulation components. These may vary both in terms of relative importance to the overall model and in terms of their internal structures. In evaluating these models we will distinguish between the component parts, and discuss the importance of each part to the overall model.

CLUG Community Land Use Game

The first model of this type which we will briefly describe is the Community Land Use Game (CLUG) model developed by Feldt (1972). The CLUG model is a manually-operated table game in which the gaming model is the dominant aspect of the exercise. In CLUG the players develop a community on a pre-existing grid system of roads. Players in the game represent industrialists, businessmen, homebuilders and residents simultaneously. Money enters the system from the sale of goods by industry to the outside. Money leaves the system for the purchase of land, for the construction of buildings, and for the purchase, through taxes, of municipal services. Players buy and sell land, construct buildings, negotiate trade and employment contracts and vote on proposed extensions of municipal services. The simulation part of the model operates to assess and reassess property and to condemn buildings. The assessment algorithm is based on the value of nearby properties. The condemnation procedure is stochastic and takes into account the age of the buildings and the players' renovation expenditures on their properties.

The objectives of the CLUG model are educational and emphasize the forces that determine the process of urban growth. Thus, the gaming model is designed to deal with the important factors that determine locational patterns. The basic model is explicitly not oriented toward objectives of prediction, optimization or formulating public policy, though the last of these has been involved in the design of experiments to supplement the basic game. Rather than evaluate the manually operated CLUG model, we will move on to a discussion of a computer based derivative CITY-RIVER BASIN MODEL.

RIVER BASIN MODEL

We turn now to two simulation-gaming models, RIVER BASIN (U. S. Environmental Protection Agency, 1972a and House and Patterson 1972) and CITY MODEL (Envirometrics, 1971b) which are evolutionary developments that started with the CLUG concept. The CITY MODEL was made operational in 1970 and was based on an extension of *City I* which, in turn, was based on the CLUG format involving expansion of the gaming component and very extensive elaboration of the simulation component. The model requires 30-60 players and easy access to computer facilities for operation of the simulation package. The RIVER BASIN MODEL is an expanded version of CITY MODEL with the addition of a water resources component to both the simulation and gaming phases. The description which follows will be for **RIVER BASIN** but will apply in many respects to the **CITY MODEL** as well.

The play of **RIVER BASIN** is centered around the development of a community in a geographical area which is divided into a 25 x 25 unit grid. Unlike CLUG, the grid lines in **RIVER BASIN** do not automatically represent roads. These lines are road beds on which future construction may take place at the discretion of the transportation department of the government in the game. The length of a grid side is 2.5 miles. The players in the game become decision makers in one or more of three sectors: economic, social, and governmental. Within each sector further disaggregation of roles is defined. Economic decision makers for instance, control the construction and operation of business activity: in four broad groupings: basic industry (both manufacturing and non-manufacturing), service industry, residential development, and agricultural activity. Within each of these there is in turn further disaggregation into specific economic activities (11 types of manufacturing, 4 types of services, etc.). The social sector decision makers represent and make decisions for the population of the area. The population is divided into three classes which are differentiated on the basis of income, education, voting status, family size, and worker status. The government sector decision makers represent the main political and bureaucratic decision makers in the system. There is an elected mayor (chairman), and there are also department heads concerned with tax assessment, water supply, utilities, municipal services (fire and police), planning and zoning, schools, highways, and public transit. The many players within each sector must interact among themselves and with players in the other sectors in seeking to achieve their self established objectives.

The gaming phase of a model run is completed with the filling out of decision forms by the players. These forms are converted into input for the simulation part of the model which is a highly complex computer program. On the basis of the gaming decisions made by the players, this program then computes new values for such key variables in the model as population patterns, water quality, employment patterns, shopping patterns, use of transportation routes, allocation of students to schools, and the distribution of commercial activity between areas of the city.

The model is more disaggregated than in the case of CLUG and the simulation component plays a major role. In CLUG, players represent several roles simultaneously, there are relatively few numbers of roles, and there is relatively little variety of choice in decisions. In **RIVER BASIN** players usually concentrate their efforts on one or two roles, there are a wide variety of roles, and the decision choices are very diverse. Furthermore in the CLUG model most of the calculations and decisions are made by the players themselves. In contrast, in **RIVER BASIN** all of the significant allocational decisions and computations are handled by the computer. It should be noted, however, that substantial aggregation remains in the model. For example each population unit represents 500 people, and each economic activity represents the average characteristics of firms in a general type of business activity. This aggregation is necessary both to keep the gaming phase manageable and also to reduce computing costs in the simulation phase.

The **RIVER BASIN MODEL** was designed to achieve the educational objective discussed earlier and to enhance our understanding of the basic forces which influence urban development. It is not concerned with issues of prediction or optimization.

In evaluating the model we will consider the gaming component first. The model includes a large number of variables that are generally considered relevant to a systemic view of urban areas. It was developed, according to its designers, from an analysis of the major activities which take place in an urban area and the gaming component was designed to coincide with a general model of urban activities. The fact that the present formulation of the model represents the latest step in a long series of model development by the designing group must be credited for some of this inclusiveness.

The gaming phase of the model appears to be consistent with current theory to the extent that current theory discusses the type and nature of the roles involved in urban processes. Furthermore, the game component of the model, to the extent that it seeks to include important aspects of the urban system which have not been developed into theoretical models (e.g., the choice between recreation and politics, or corruption in political and economic activity), plays a crucial role in reducing the possible use of inconsistent or invalid theory in the simulation model. Where good theory does not exist, interactions and decisions are gamed, whereas when good theory does exist, it is often incorporated into the simulation component of the model.

As an educational device the **RIVER BASIN MODEL** may be nearly as complex as it can be and still achieve its objective. Experiences with the game and its predecessor **CITY MODEL** suggest that the number of players, the complexity of the decisions facing them, and the amount of information flow are nearly at their maximum level for effective management of the game by both the players and the operators (see U. S. Environmental Protection Agency, 1971b).

With regard to the model's consistency with data, the designers suggest that the design and specification of the model itself was not initially directly dependent on specific data. Operations of the model have, until recently, depended on hypothetical starting data.

APEX—Air Pollution Exercise

The final model which we will consider in this general area of urban simulation-gaming is the **APEX** model (U. S. Environmental Protection Agency, 1972b). The **APEX** model, which was first made operational in 1968–69, is a successor to the **METRO** model and the **METROPOLIS** model, all of which were developed under the leadership of Richard Duke at the University of Michigan's Environmental Simulation Laboratory. The historical roots of **APEX** make it quite different from the **CLUG-RIVER BASIN** type of models. This model simulates the conditions of the Lansing, Michigan area and the roles that have been significant in the development of that region. Unlike **RIVER BASIN**, **APEX** is composed of two operational components, a gaming phase and a simulation phase.

The gaming sequence involves players in one of five general roles: politicians, planners, industrialists, commercial developers, and air pollution control officers. Within these sectors there is additional differentiation between roles in the city and in the surrounding county and between types of industry and development. It is noteworthy that households are not included in the gamed component of the model. Their inclusion in the simulation component reflects the historical development of the model. Whereas the **RIVER BASIN** designers sought to simulate the activities of the urban system, the designers of **APEX** and its predecessors have sought to model the roles of the "principal" actors or decision makers. This emphasis on identifiable decision makers has meant that this model deals more with the environmental context of given roles than with the nature of the system per se.

The emphasis in the game component on specific roles rather than activities has meant that the simulation component simulates some processes that are gamed in **RIVER BASIN**. The simulation component is comprised of a number of interacting submodels which (1) determine growth in export industry employment, (2) allocate local service sector employment within the different parts of the urban area, (3) simulate voter response to bond and tax propositions put before them, (4) determine the results of political contests between incumbents, as gamed, and their simulated opponents, and (5) determine the amount of air pollution in the urban area.

Besides these main submodels there are a series of major subroutines which perform important model functions. These include routines dealing with capital plant, land value, prices, zoning, taxes, entering firms, purchasing and accounting, accessibility, non-white population, and the news media.

The APEX model may be characterized as somewhat more aggregated than the RIVER BASIN MODEL. The relative aggregation concerns the number of locations (29 analysis areas) into which activities are allocated by the simulation model (as compared to 625 in RIVER BASIN) and the number of role types involved. The APEX model has very detailed roles in the government and business sectors but leaves the household sector to the simulation model.

The APEX model, like the others discussed in this section, was designed to meet educational objectives and to improve our understanding of the urban development process. Specifically, the current version of the APEX model was designed to meet the needs of the Air Pollution Control Institute and its training program. The exercise is now a part of the Institute's program to develop administrators to take active roles in air pollution control projects in our urban areas. Thus, the educational objective of the model which was explicit in its design is also explicit in its use.

APEX is the last of the specific simulation-gaming models which we shall look at in detail. Before leaving this general class of models, however, a few general comments may be useful. As distinct from some of the other models we have considered, these simulation-gaming models seek to represent a systemic or holistic concept of urban areas. As such they represent an important contribution to the study of urban affairs. They are built on the theory that some human interactions cannot be modeled explicitly and must therefore be left to humans to perform. The simulation-gaming format allows the designer to combine these un-modeled interactions with accepted theory in the development of a more complete model. We may learn more about the systems from these models and we may also study the content of these human interactions. These factors give these models considerable potential for experimentation.

V. CONCLUDING COMMENTS

This completes our survey of urban modeling. There has been no attempt to survey all of the many urban models which have been created over the past two decades, but we have discussed a number of important models of different types which are representative of some of the major trends in urban model design and here indicated to the reader that there is a large literature of surveys and evaluations. It may now be useful to conclude this chapter by taking a brief overall look at the degree to which these models have been able to accomplish their objectives and at their usefulness for policy makers.

In general, it would appear that substantial progress has been made in urban modeling in the past 20 years and that a great deal has been learned in that time both about cities and about techniques for modeling them. Nevertheless, it is clear that urban modeling efforts have not been completely successful in accomplishing all of their objectives. So far, for instance, no single model has fully succeeded in achieving both rigorous consistency with theory and also close calibration with the data from an individual city.

The empirical land use models perhaps go the farthest in the direction of close calibration with specific urban data. But they have often had to make compromises with theory, and there have been difficulties in making them operational. The Forrester model and some of the urban simulation-gaming models suggest interesting hypotheses about the dynamic relationships present in an urban area, but they, too, have not been fully coordinated both with all of the existing theoretical literature and also with detailed and accurate data.

We conclude this chapter by first citing critics of planning models in general and urban models in particular, and then by offering our view of the present, proper role of such models.

Lee in his article "Requiem for Large-Scale Models" in the *Journal for the American Institute for Planners*, (Lee, 1973) offers the following critical assessment:

> "The task in this paper is to evaluate, in some detail, the fundamental flaws in attempts to construct and use large models and to examine the planning context in which the models, like dinosaurs, collapsed rather than evolved. The conclusions can be summarized in three points: (1) In general, none of the goals held out for large-scale models have been achieved, and there is little reason to expect anything different in the future; (2) For each objective offered as a reason for building a model, there is either a better way of achieving the objective (more information at less cost) or a better objective (a more socially useful question to ask); (3) Methods for long-range planning—whether they are called comprehensive planning, large-scale systems simulation, or something else—need to change drastically if planners expect to have any influence on the long run."

Lee directs criticism toward the current urban modeling efforts related to Forrester's Urban Dynamics model, the NBER model, and the Projective Land Use Model (PLUM). He sees little sign of success in either the past or current urban modeling efforts.

Brewer (1973) has a more lengthy critique that is based upon four types of appraisal (theoretical, technical, ethical, and pragmatic). He deals thoroughly with the two community renewal program models (San Francisco and Pittsburgh). His focus is on the decision-making process in these two complex urban settings and the attempts to use large scale computer simulation models for planning and development. Brewer's book could benefit a public official contemplating the construction or use of a large scale computer model. It provides the background and evaluation criteria necessary to evaluate the intended model-building process and product.

In the light of these observations about urban models, then, it may be of interest to briefly consider their general usefulness in policy making. With regard to this question, there is a key distinction to be made between the use of models as a *tool for policy making* on the one hand and as a *replacement for policy makers' judgment* on the other. It is clear from the above analysis, that urban models are not now developed to the point where they can be mechanically used as a substitute for a policy maker's judgment in urban decision making. We do not have a model which will simply allow the public official to "plug in the numbers" and get back the correct decision with regard to a housing or transportation or zoning decision. None of the current models can do this, because none of them can fully capture the extreme complexity of the urban public official's decision-making environment.

Urban models can, however, be extremely useful as a *tool* or an *aid* to urban policy making. They are not intended to *replace* the urban policy maker's judgment, but they can be extremely useful in *improving* his judgment. The empirical land use models, for instance, can be a valuable aid in helping the urban planner assemble and manipulate a large amount of data about urban land use in a systematic and consistent way. And the Forrester and simulation-gaming models can help alert the policy official to the complex patterns of interrelationships in the urban area and to the possibility of counterintuitive results from his policy decisions.

Thus urban models can be potentially useful. But ultimately it is the policy maker's task to use both the models and his own judgment in attempting to arrive to correct policy decisions. The task of the urban model builder in the future must be to continue to find ways to integrate both theory and data into models which will be of further use in helping the policy maker with his job.

References

Almendinger, V. V., *Topics in the Regional Growth Model: I*, Paper No. 4, Penn-Jersey Transportation Study, Philadelphia, 1961.

Anundsen, Kristin, and Nilo Londgren, "Can You Put a Simulation Model to Work in a Real Community?", *Innovation*, 29, March 1972.

Applied Decision Systems, Inc., *Prototype for the Regional System Simulation and Control Model*, Prepared for the Houston-Galveston Area Council, January 1972.

Arthur D. Little, Inc., *Model of San Francisco Housing Market*, San Francisco Community Renewal Program Technical Paper No. 8, January 1966.

Batty, M., *Recent Developments in Land Use Modeling: A Review of British Research*, Urban Systems Research Unit, Department of Geography, University of Reading, Reading, England, September 1970a.

Batty, M., "An Activity Allocation Model for the Nottinghamshire-Derbyshire Subregion," *Regional Studies*, Vol. 4, No. 3, 1970b, pp. 307–332.

Batty, M., "Models and Projections of the Space-Economy," *Town Planning Review*, Vol. 41, No. 2, pp. 121–147, April 1970c.

Berry, Brian J. L., "The Retail Component of the Urban Model," *Journal of the American Institute of Planners:* Vol. XXXI, No. 2, May 1965.

Birch, David et. al., *The New Haven Laboratory: A Testbed for Planning*, Harvard University, Cambridge, April 1973.

Black, J., "Some Retail Sales Models," A paper presented at the Urban Studies Conference, United Kingdom, 1966.

Brewer, Garry D., *The Politician, The Bureaucrat and The Consultant: A Critique of Urban Problem Solving*, Basic Books, New York, 1973.

Brown, H. J. et al., *Empirical Models of Urban Land Use: Suggestions on Research Objectives and Organization*, National Bureau of Economic Research, 1972.

Center for Real Estate and Urban Economics, Jobs, People, and Land: Bay Area Simulation Study (BASS), Special Report No. 6, University of California, Berkeley, 1968.

Chapin, F. Stuart, Jr., "A Model for Simulating Residential Development," *Journal of the American Institute of Planners*, Vol. XXXI, No. 2, May 1965, pp. 120–125.

Chicago Area Transportation Study, *Final Report*, Vol. 2 and 3, Chicago, 1960.

Crecine, John P., *A Time-Oriented Metropolitan Model for Spatial Location*, CRP Technical Bulletin No. 6, Pittsburgh, Department of City Planning, January 1964.

Cripps, E. L. and D. H. S. Foot, "A Land-Use Model for Sub-Regional Planning, *Regional Studies*, Vol. 3, pp. 243–68, 1969.

Donnelly, Thomas F., Stuart Chapin, and Shirley F. Weiss, *A Probabilistic Model for Residential Growth*, Urban Studies Research Monograph, The Institute for Research Social Science, Chapel Hill, University of North Carolina Press, 1964.

Duke, Richard and Richard L. Meier, "Gaming Simulation for Urban Planning," *Journal of the American Institute of Planners*, Vol. XXXII, No. 1, January 1966, pp. 3–17.

Echenique, M. et al., "Development of a Model of a Town," Working Paper 26, Cambridge, England: Center for Land Use and Built Form Studies, pp. 59, 1969.

Envirometrics, Inc., *State-of-the-Art of Urban Gaming Models*, National Technical Information Service, Springfield, Virginia, p. 126, 1971a.

Envirometrics, Inc., *City Model Players' Manual* (Director's Guide and Operator's Manual), Technical Analysis Division, National Bureau of Standards, U. S. Department of Commerce, Washington, D. C., 1971b.

Feldt, Allan, *CLUG: Community Land Use Game*, The Free Press, New York, 1972.

Forrester, Jay, *Urban Dynamics*, MIT Press, Cambridge, Massachusetts, 1969.

Garin, Robert A., "Research Note: A Matrix Formulation of the Lowry Model for Intrametropolitan Activity Allocation,' *Journal of the American Institute of Planners*, pp. 361–366, November 1966.

Goldberg, M., "Industrial Location Model for the San Francisco Bay Area," A paper presented at the Western Regional Science Association, Las Vegas, Nevada, 1967.

Goldner, William, "The Lowry Model Heritage," *Journal of the American Institute of Planners*, Vol. XXXVII, No. 2 March 1971.

Goldstein, Gerald and Leon Moses, "A Survey of Urban Economics," *Journal of Economic Literature*, Vol. XI, No. 2, June 1973.

Hamilton, H. R., S. E. Goldstone, J. W. Milliman, A. L. Pugh, III, E. B. Roberts and A. Zellner, *Systems Simulation for Regional Analysis: An Application to River-Basin Planning*, Institute of Technology Press, Cambridge, Massachusetts, 1969.

Hansen, Walter G., "How Accessibility Shapes Land Use," *Journal of the American Institute of Planners*, Vol. XXV, No. 2, pp. 73–76.

Harris, Britton, "Urban Development Models: New Tools for Planning," *Journal of the American Institute of Planners*, Vol. XXXI, No. 2, May 1965.

Harris, Britton, "Quantitative Models of Urban Development," *Issues in Urban Economics*, Johns Hopkins Press, 1968.

Hemmens, George, *Urban Development Models*, Highway Research Board Special Report 97, Washington, D. C., 1968.

Hemmens, George C., "Urban Development Modeling," A paper presented at the Program of Policy Studies in Science and Technology, George Washington University, Washington, D. C., 1970.

Hendricks, Francis, "Problems of Large Scale Simulations: Some Lessons From the Design of GSPIA. A paper presented at the Annual Conference of the Urban and Regional Information Systems Association and Concurrent Government Management Information Sciences Conference, Information Research for an Urban Society, San Francisco, August 1972.

Herbert, John D. and Benjamin H. Stevens, "A Model for the Distribution of Residential Activity in Urban Areas," *Journal of Regional Science*, 1960.

Hill, Donald, "A Growth Allocation Model for the Boston Region," *Journal of the American Institute of Planners*, Vol. XXXI, No. 2, 1965.

House, Peter W. and Philip D. Patterson, editors, *An Environmental Laboratory for the Social Sciences*, U. S. Environmental Protection Agency, Washington, D. C., 1972.

Huff, D. L., "Determination of Intraurban Retail Trade Areas," Mimeo., Real Estate Research Program, University of California, Los Angeles, California, 1962.

Ingram, G. F., J. F. Kain, and J. R. Ginn, *The Detroit Prototype of the NBER Urban Simulation Model*, New York, 1972.

Kilbridge, Maurice et al., *Urban Analysis*, Graduate School of Business Administration, Harvard University, Boston, Massachusetts, 1970.

Lakshmanan, T. R. and Walter G. Hansen, "A Retail Market Potential Model," *Journal of the American Institute of Planners*, Vol. 31, No. 2, pp. 134–143, May 1965.

Lakshmanan, T. R., "A Model for Allocating Urban Activities in a State," *Socio-Economic Planning Sciences*, Vol. 1, No. 3, 1968.

Lamb, Donald, *Research of Existing Land Use Models*, Southwestern Pennsylvania Regional Planning Commission, Pittsburgh, March 1967.

Lathrop, George T., John R. Hamburg, and G. Frederick Young, *An Opportunity-Accessibility Model for Allocating Regional Growth*, Albany, New York, State Department of Public Works, Subdivision of Transportation Planning and Programming, January 1965.

Lee, Douglass, *Models and Techniques for Urban Planning*, Cornell Aeronautical Laboratory, Buffalo, New York, 1968.

Lee, Douglass, "Requiem for Large-Scale Models," *Journal of the American Institute of Planners*, Vol. XXXIX, No. 3, May 1973.

Levin, Melvin R. and Norman A. Abend, "The Penn-Jersey Transportation Study," *Bureaucrats in Collision: Case Studies in Area Transportation Planning*, MIT Press, Cambridge, Massachusetts, 1971.

Lowry, Ira, *A Model of Metropolis*, RM-4035-R6, The RAND Corporation, 1964.

Lowry, Ira, "A Short Course in Model Design," *Journal of the American Institute of Planners*, Vol. XXXI, No. 2, May 1965.

Lowry, Ira, "Seven Models of Urban Development: A Structural Comparison," *Urban Development Models*, edited by George C. Hemmans, Highway Research Board Special Report 97, Washington, D. C., 1968.

McLoughlin, J. B., C. K. Nix and D. H. S. Foot, *Regional Shopping Centers: A Planning Report in Northwest England, Part 2: A Retail Shopping Model*, Department of Town and Country Planning, University of Massachusetts, 1966.

Meyers, Charles R., Jr., *Regional Modeling Abstracts*, Regional Environmental System Analysis Program, Oak Ridge National Laboratory, Oak Ridge, Tennessee, Vol. I–V, June 1973.

Miller, Roy I. et al., *Mathematical Models for Model Cities Evaluation: A Feasibility Study*, The Institute for Creative Studies for the National Bureau of Standards, U. S. Department of Commerce, Washington, D. C., 1970.

Music, B. and B. Barber, "Description of Alternative Patterns and Comparison of Lowry and 'Hand' Forecasts," *Spatial Policies for Regional Development*, Technical Report 11, American-Yugoslav Project, 1970.

Nagelberg, Mark and D. Little, *Simulation of Urban Systems—A Selected Bibliography*, Institute for the Future, Middletown, Connecticut, January 1970.

Nagelberg, Mark and D. Little, *Selected Urban Simulations and Games*, Institute for the Future, Middletown, Connecticut, March 1970.

Putman, Stephen H., *Industrial Location Model (INLNP)*, Technical Bulletin No. 5, CONSAD Research Corporation, December 1963.

Rhodes, T. and R. Whitaker, "Forecasting Shopping Demand," *Journal of the Town Planning Institute*, Vol. 53, pp. 188–192, 1967.

Robinson, Ira, Harry B. Wolfe, and Robert L. Barringer, "A Simulation Model for Renewal Programming," *Journal of the American Institute of Planners*, Vol. 31, No. 2, pp. 126–133, May 1965.

Rose, J., *Intraregional Industrial Location Theory and Methodology*, A paper drafted for the Regional Planning Council, 1967.

Schlager, Kenneth J., "A Land Use Plan Design Model," *Journal of the American Institute of Planners*, Vol. 31, No. 2, pp. 103–110, May 1965.

Seidman, David R., *Report on the Activities Allocation Model*, PJ Paper No. 22, Philadelphia, Pennsylvania, Penn Jersey Transportation Study, November 17, 1964.

Seidman, D. R., "The Construction of an Urban Growth Model," *DVRPC;* Philadelphia, Pennsylvania, p. 229, 1969.

Steger, Wilbur A., "The Pittsburgh Urban Renewal Simulation Model," *Journal of the American Institute of Planners*, Vol. 31, No. 2, pp. 144–149, May 1965.

Sweet, David C., editor, *Models of Urban Structure*, Lexington Books, D. C. Heath and Company, Lexington, Massachusetts, 1972.

Traffic Research Corporation, *Review of Existing Land Use Forecasting Techniques*, Presented to the Boston Research Planning Project, 1963.

U. S. Environmental Protection Agency, *The River Basin Model* (14 Volumes), Water Pollution Control Research Series, 16110 FRU12/71, U. S. Government Printing Office, Washington, D. C., 1972a.

U. S. Environmental Protection Agency, *APEX: Air Pollution Exercise* (21 Volumes), Office of Manpower Development, Office of Air Programs, U. S. Environmental Protection Agency, Washington, D. C., 1972b.

Voorhees, Alan M., editor, "Land Use and Traffic Models: A Progress Report," Special Issue of the *Journal of the American Institute of Planners*, Vol. 25, No. 2, May 1959.

Wilson, A. G., "Development of Some Elementary Residential Location Models," *Journal of Regional Science*, Vol. 9, pp. 337–385.

Chapter 7

Models in Transportation

By

Kenneth W. Webb
(with contributions by)
Frank L. Spielberg and Peter S. Loubal

SUMMARY 203

I. INTRODUCTION 204
- Decision Making Agencies 206
- General Types of Models 208

II. URBAN AREA TRANSPORTATION PLANNING 208
- The Urban Transportation Problem 208
- The Need for Urban Transportation Planning 209
- Transportation Planning Techniques 209
- Models for Urban Transportation Planning 210
- Overview 210
- Land Use Model 211
- Bay Area Transportation Land Use Model 213
- EMPIRIC Land Use Model 213
- Trip Generation Models 214
- Zonal Interchange Model 215
- Trip End Modal Split Models 216
- Trip Interchange Modal Split Models 216
- Network Assignment Models 217
- Deficiencies of the Traditional Approach 217
- Recent Developments and New Directions 218

III. INTER-URBAN TRANSPORTATION PLANNING 220

IV. NATIONAL TRANSPORTATION PLANNING 222
- A Transportation Assessment Model 223
- Airport Investment Model 225
- An Aircraft Route Effectiveness Model 226
- Summary 227

REFERENCES 227

Models in Transportation

SUMMARY

Since 1950 the use of scientific modeling as an aid to transportation decisions has developed to the point where the total effort exceeds modeling use in all other areas except the military. This chapter contains descriptions of the transportation decision-making problems and the models that have been employed to solve them. Transportation decision making interrelates to practically all social and economic domains of local and national government. This causes the decision making to be particularly difficult and the models to be rather complex. A new freeway that is in the proposal stage must be evaluated not only with regard to the effect on traffic but also with regard to secondary effects on such matters as land development relieving unemployment and social access. Negative effects such as neighborhood disruption, pollution and aesthetics must also receive attention. Building models that will assess this large number of characteristics is, to say the least, difficult. However, the rewards in terms of more accurate planning judgments make the expense and effort worthwhile.

Transportation decision making is in transition, bringing about new facets in the modeling approaches. There is a general shift in governmental attention away from pure analysis of trips on airplanes and in automobiles, to analysis of socio-economic effects on the citizens who are riders or in the pathways. Citizen groups now have political power to prevent proposed highways and airports because of infringements on their rights. The recent change in the use of highway funds from strictly highway construction to the use for mass transit is another change in terms of technological approaches.

The variety of models that have been developed and used on transportation problems is quite extensive. They range from multi-million dollar regional studies to small models of alternative designs for intersections in a city. The models range over many facets of transportation: school bus scheduling, railroad scheduling, highway location, airport design, maintenance of equipment, crew scheduling, trash truck routing are only a few examples.

This chapter presents the larger decision problems and the related models. It describes modeling approaches for problems of a national or large local area importance. Models of smaller impact, such as intersection or traffic signal system design are not problems in the large and are not covered.

The first section of this chapter describes urban transportation planning which is in a real state of transition. The movement of citizens to cities and the increasing attention to the variety of needs of citizens is bringing an increasing importance to the planning function, which is for the most part concerned with investments and their effects.

The models here can be classified as follows:

- Macro-Level—dealing with major traffic corridors
- Messo-Level—generally dealing with route location studies
- Micro-Level—generally design planning such as intersections

At the macro-level, modeling is composed of a set of submodels. First, a

Preceding page blank

land use model is constructed to represent the land area in the region, the uses of the land, and the transportation network. Models are then developed for generating trips that the citizens and freight haulers might make, distributing these trips on the network, and making choices of modes of travel and the selection of the routes that each trip maker might take. These models are described in some detail in the chapter to illustrate their complexity and comprehensiveness.

The next section discusses a megalopolis model developed to analyze the transportation problems of the Northeast section of the United States. The Federal government in the early 1960 period recognized that improvements in this geographic area were necessary because of the poor condition of the railroads and the congestion of the highways. New technology such as tube trains and high speed surface trains were tested on the model. It was also found necessary to review investment options and changes to public policy which could be evaluated by the large model. The Northeast Corridor Model was actually composed of many submodels covering both freight and passenger traffic. Demand was estimated under several scenarios, followed by mode and trip allocation to a network model. Demand estimation was followed by a model to assess the impact of proposed networks incorporating the innovations.

The third section of this chapter is concerned with national transportation planning. The Federal Department of Transportation plans for movement of goods and people on a national scale and has a mission to disburse monies to State and local governments for investment in facilities. The planning role is to determine the best disbursement on a national basis to meet the program needs and priorities. Additionally, the Department is responsible for determining new legislation to meet the new special needs.

In response to the evaluation needs for disbursement of Federal monies to the States and local governments, a national model was developed which assessed the highways, arterials, connectors and local roads of the nation. Seventy-four parameters such as accident rates, pollution, speeds, and amount of traffic are input to the model. This section describes the models used to evaluate this enormous amount of data and made assessment of the best allocation of monies throughout the nation. The national model results do not correlate well with the needs for investment as seen by the localities. This reflects the lack of conformity on objectives between the Federal and local governments.

This section also describes a model used for the development of airports in the United States. Airports exemplify the complexities of public utility investment. Federal bureaus regulate the air carrier industry while State and local commissions make the airport capacity decisions. The public demand for service is related to the capacity of the system and the behavior of the carriers. The model described relates the capacity and the cost allocation decisions that must be made by the Federal government.

At a more detailed level, the Federal government has decision problems in evaluating airline operations for determinations on policies concerning fares and user charges, route awards, requests for service charges, mergers and intercarrier agreements, new terminal decisions, limits on airport operations and effects of new aircraft. A model is described which develops a method to approximate the mix of planes, routes and schedules and terminal facilities that will satisfy intercity passenger and cargo demand at minimum social and economic costs.

The following Summary Table lists the models that are described in this chapter on transportation.

I. INTRODUCTION

It would be difficult to overstate the importance of transportation in shaping

TRANSPORTATION

Models Discussed in Chapter

Summary Table

Model Decision Area	General Type	Important Characteristics
Urban Area		
Land Use Model	Trend or Regression Analysis	The urban region is divided into zones and employment and population and determined for each
Trip Generation model	Regression	The frequency of origins or destinations of trips in each zone is determined by land use and socio-economic factors
Zonal Interchange model	Gravity Model	Based on data from origin and destination surveys, the origins by zone are distributed by a model as destinations to zones
Trip End Modal Split Models	Trend Analysis	The origin or destination statistics are divided among modes according to the factors of trip, trip maker and mode
Trip Interchange Modal Split Models	Linear Regression	The traffic between zones is allocated to the available modes of transportation according to factors of trip, trip maker and mode
Network Assignment Model	Moore Algorithm	All trips are assigned to networks based on minimal time or minimal cost to the trip maker
Inter-Urban Transportation Planning		
Northeast Corridor Model	Simulation and a variety of Optimization techniques	A model for testing innovations in the North East megalopolis for freight and passenger systems
National Transportation Planning		
Transportation Assessment Model	Ranking Analysis Cost Benefit Analysis	A national model to determine the best disbursement of Federal funds to States and localities
Airport Investment Model	Dynamic Programming	The model quantifies the benefits received and the capacity required by various airport user groups
Airport Route Effectiveness Model	Gravity Model, Linear Programming	This model is used to analyse the operational characteristics and management characteristics of airlines and airports

not just the overall commercial and social patterns of the United States, but the individual life style of each citizen. The tremendous benefits of increasingly faster and cheaper transportation were achieved by an essentially spontaneous development of transportation means—whether on waterways, railroads, highways or in the air. The attendant problems of uncoordinated growth and planning—railroad bankruptcies, decline of public transit—require solutions based on a more rational approach to both the industry's and public's often conflicting concerns. In addition, transportation decision-makers are also being confronted with concomitant problems of pollution, inefficient land use, and energy shortages. Resolution of the interrelated local, State and Federal transportation problems requires an increasing reliance on analysis, modeling and other procedures for alternative evaluation and selection.

Historically, transportation problems were among the earliest to be studied by the emerging modern management sciences, e.g. accessibility and facility location—the Weber problem [76], least-cost transportation connections—the Hitchcock problem [59], shortest paths through networks—the Moore Algorithm [62]. These early efforts and more recent developments now form a valuable set of transportation decision-making aids. These include the Urban Planning Package of the Federal Highway Administration [57] for highway planning, and the multimodal UMTA Transportation Planning System of the Urban Mass Transportation Administration [56], [73], [74] for private and public transport planning.

The need for such aids grew out of their incipient use in such major multimillion dollar regional ground transportation studies as the Chicago Area Transportation Study—CATS; Tri-State Transportation Study—New York City; Pittsburgh Area Transportation Study —PATS; Penn-Jersey Transportation Study—Philadelphia; and the Bay Area Transportation Study—BATS. The latter study was concerned with the development of a regional ground transportation plan for the nine county San Francisco area. Large amounts of information were collected describing the economic activities, population characteristics, transportation facilities and land use characteristics. Data sources included public record, household interviews and private agency studies. The data were input to a large regional growth model which is described later in this report.

Another example of the scope of large transportation studies was a modeling effort for the State of California. The objective of this effort was to plan a transportation complex for the long range (50 years) to meet the needs of the state for growth. The study team examined all possible transportation modes—tube trains, air cushion equipment, sky buses, electric automobiles and many others. The complex interrelated aspects of population, land use, economy and technology with regional, interurban and urban transportation needs were examined using the set of models depicted on Chart I.

Many models exist for special problems in local transportation, of much smaller dimensions. For instance, the problems of scheduling buses has produced computer models to bring more efficiency to the operations [41], [35], [36]. The experiments in Dial-a-Ride bus systems were modeled for initial testing of the system [39]. Models have been developed for a multiude of design problems for intersections, ramp entrances, traffic signal control systems, school bus scheduling, etc. However, these problems and their models will not be discussed in this report because they have little effect on overall planning for urban areas.

In the last half dozen years the transportation decision processes have become more concerned with the effects of transportation systems on the users and the people in the pathways. The effects of noise, air pollution, disruption of neighborhoods, hardship of being moved, effects on small business and even church memberships, and many other effects are being studied and modeled to a greater degree. (See [1], [2], [3], [5] and [6]).

Decision Making Agencies

The decision making agencies in the field of transportation that have used models are numerous. Some of the most important are the following:

- City, County, and State departments that have problems in the structuring of road nets to serve the needs of comprehensive planning.
- Regional and inter-State planning bodies with decision making problems of organizing highway and road nets that serve the needs of the larger-than-State areas for growth and control.
- Federal agencies that are responsible for the disbursement of funds to states for the construction and extension of networks to serve the needs of the country.
- Federal, State and local agencies responsible for solving problems of

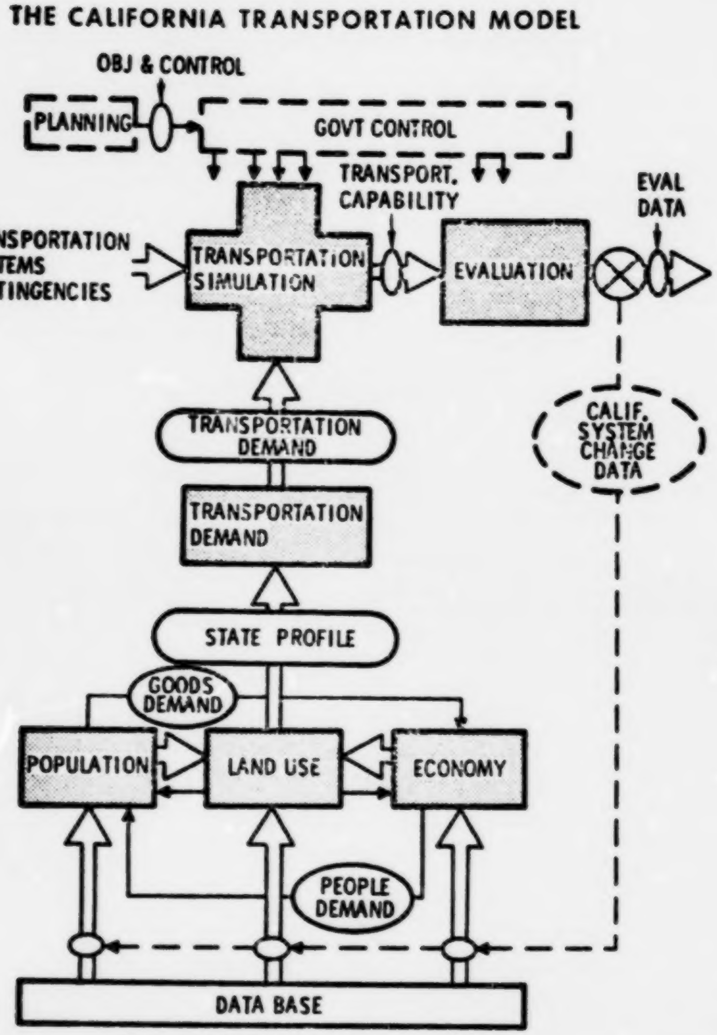

THE CALIFORNIA TRANSPORTATION MODEL
OBJ & CONTROL
PLANNING
GOVT CONTROL
TRANSPORT. CAPABILITY
EVAL DATA
TRANSPORTATION SYSTEMS CONTINGENCIES
TRANSPORTATION SIMULATION
EVALUATION
CALIF. SYSTEM CHANGE DATA
TRANSPORTATION DEMAND
TRANSPORTATION DEMAND
STATE PROFILE
GOODS DEMAND
POPULATION
LAND USE
ECONOMY
PEOPLE DEMAND
DATA BASE

CHART I—*The California Transportation Model*

Source: *Systems Technology Applied to Social and Community Problems*, June 1969. Committee on Labor and Public Welfare, United States Senate.

highway intersection design, traffic control methods, parking needs and other requirements of smaller scale engineering.

- Regulatory bodies such as the Interstate Commerce Commission (and some industry groups) that fix rates that transportation systems are allowed to charge.
- Federal, State, and sometimes local agencies that regulate air traffic, operations of airports and control the flights.
- Agencies at all levels of government that regulate railroads and shipping companies.
- Transportation industries that make decisions on proposed routes

> schedules, crew scheduling, facilities, maintenance and many other operational matters.

General Types of Models

As one can imagine, with the large number of agencies involved in decision making, we will find wide diversity in transportation models. Generally these fall into three main categories, although there are exceptions.

(1) Transportation land use models, which for an area under study, characterizes the relationships among employment, housing, industry, and transportation. The aims in these models are to determine the transportation network that promotes growth, employment and other desirable characteristics. Transportation land use models will be described in detail in this chapter because of their importance to large area decision making.

(2) Linear programming, dynamic programming and other non-probabilistic models are used generally to solve decision problems where an optimum result is needed and the problem can be characterized by sets of equations. The use of linear programming in the solution of least cost transportation routes for goods going from several origins to several destinations is classic in the field. An application of dynamic programming to determine a sequence of airport investments will be described in detail in this paper.

(3) Simulation has been used extensively because it provides for the build up of queues of autos, airplanes, trains, passengers, school buses, public buses and other objects in motion and being serviced at points in the trip. The modeling objective is usually to determine the system that minimizes numbers in quences or the time in queues. For instance, the Long Island Railroad, which has an unbeaten queuing history, recently simulated its entire operation to determine schedules, number of equipments required, locations for new facilities, and personnel needs. The objective was to find the system that minimized total trip times, waiting times and random breakdowns. Simulation and queuing theory models are used for solving operational problems for the most part. They will not be treated here because their relative impact on governmental decision making is lower than deterministic models.

The sections which follow will present only a few of the many models used in governmental decision making. Selections of transportation land use models will be reviewed in detail because of their significance in local and regional planning. Some of the representative models used at the Federal government level will also be reviewed in detail as they have strong national impact.

II. URBAN AREA TRANSPORTATION PLANNING

The Urban Transportation Problem

The contrast between the excellent service offered by the motor vehicle in areas or during time periods when it can be provided with adequate space on the highways and sufficient parking, and the situation encountered by most drivers in metropolitan areas during the rush hours is too well known to require lengthy elaboration. Excessive use of the *private automobile* can be characterized by the following negative factors.

> *Congestion*—massive efforts of highway engineers have not eliminated the frustration, time losses and costs stemming from the incapability of street and parking facilities to cope with ever-increasing peak demands.
>
> *Inefficient Land Use*—the amount of land needed to serve vehicles in inner core areas has been increasing to a degree where the scarcity of land available for the ultimate trip purpose (employment, shopping, recreation, etc.) leads to sprawls that further the demand for street space. This causes an upward spiraling of the problem.
>
> *Air Pollution*—other types of pollution are wasteful and unpleasant. Air pollution, mainly from motor vehicle exhaust, can become hazardous to health.
>
> *Excessive Consumption of Natural*

Resources—concern is growing that some basic natural resources, e.g., fuel oil, will be depleted before suitable substitutes can be found.

Negative Social Effects—excessive motorization reduces the level of service which can be provided by public transit to those who cannot drive—some 30 percent of the population who are too young, old or infirm to drive, and 20 percent of families who have no car available. This reduces their employment possibilities, diminishes their recreational opportunities and forces them to accept a low level of personal mobility.

Mass transit is not just geared to cope substantially more efficiently with large demands for service than the private car. A high rate of demand is a prerequisite for achieving good transit service profitably, or at least a relatively low subsidy. The post-war downward transit spiral where decreasing patronage led to increased fares and service reductions followed by further drops in patronage has now been to a large degree halted. The major question facing us all is whether and how this process can be reversed in a free-choice society.

Will transit patronage consist of only the non-drivers and those who have been forced to abandon their cars by impossible highway conditions, or can a substantial part of riders with reasonable access to transit be retaught to use it? Is its main role to deal with peak travel demands or can an acceptable level of utilization be achieved off-peak? Does conventional transit technology possess the capability to satisfy a majority of travellers, or do we have to develop novel systems to provide still greater speed, flexibility or comfort? Can a more efficient use of travel modes be achieved by a pricing mechanism?

Concurrently it must be recognized that the private automobile has had many positive social impacts and has been overwhelmingly accepted by the American public. Clearly it is not desirable to undo the work of the past seventy years. Rather the efforts in transportation planning must be directed at recognizing the undesirable impacts of auto usage and striving to develop comprehensive multi-modal systems which combine the best of all modes while minimizing the adverse impacts.

The Need for Urban Transportation Planning

A large proportion of the capital investment in any urban region is related to transport facilities. Over the past several decades, since 1948, the aggregate highway expenditure in the United States has been approximately $260 billion. These facilities are the means toward the end of providing the residents of a region with accessibility between homes, jobs, services and the like. Other means to the same end such as vastly improved communication techniques or rearrangement of land use and land activity patterns could achieve the same effect, but would likely result in a different life style than that to which we are accustomed.

For better or for worse, the majority of U. S. citizens have made the implicit decision to live in urban areas with a relatively large spacial separation of activities. Given this preference extensive transport systems become a necessity.

Transportation Planning Techniques

To provide a basis for rational decisions in investments for urban area transportation facilities the decision makers must have estimates of the probable results of a given investment related to the current and anticipated transport problems.

In an ideal situation the technician would be able to provide decision makers with future state information comparable to current state data. Thus, it could be determined if a given transport investment reduced travel costs, provided the unemployed with better access to jobs, resulted in an increased tax base, reduced air pollution, provided service to the handicapped, or responded to the host of expressed and unexpressed desires of the regional population and its component subgroups.

Unfortunately, this ideal state has not been achieved. The best that transportation planners have been able to do to date is to develop somewhat crude forecasts of some elements of the overall regional pattern. Such progress as has been made has been achieved through the use of models designed to represent urban development and travel phenomena.

Models for Urban Transportation Planning

Models in general use in urban transportation analysis have been focused on forecasts of land development patterns, the demand for travel among subareas with a region, the mode of travel and the choice of route. The results of these techniques have permitted evaluation of the usage of specific facilities with a view to required size and/or economic feasibility.

A few more specific simulation models have been developed to investigate vehicle movements on roadways, the timing of traffic signals or power consumption of rapid transit vehicles.

In general, models of urban transportation have been scaled to the types of issues to be addressed. Three broad classes may be identified.

> *Macro-level*—Dealing typically with analysis of major travel corridors only, this scale requires minimal detail. Analysis at this scale is often referred to as sketch planning.
>
> *Messo-level*—Analysis in this group are focused on specific facilities and correspond to route location studies. These models generally require a large amount of data, but precision need not be great.
>
> *Micro-level*—These models treat only a small portion of the transport system such as a transit station or a group of intersections, and are comparable to the final design stage in construction planning. A large amount of precise data is necessary for accurate results.

Overview

Modeling techniques in one form or another have been in use since the beginning of major highway studies in the early 1940's. Simple manual procedures for forecasting travel and assigning trips to routes through a network were applied in a systematic fashion, although many judgmental decisions were required.

Transportation modeling in its present form began to develop in the mid-1950's with greater availability of large scale computers. The predecessor agencies of the U. S. Department of Transportation as the primary sources of funding for transportation planning played a major role in the development of the modeling techniques in current use. In the early 1960's the Bureau of Public Roads supported development of computer programs and procedural manuals for highway planning, while the Department of Housing and Urban Development funded development of public transit planning models in the period between 1965 and 1967.

While recognizing that the transport decisions of individual travelers encompass simultaneous evaluation of the need to travel, alternative destinations, alternative modes and alternative routes, the traditional planning models have been structured to treat each element as an independent event. This structure was adopted primarily for simplicity in model development.

Five major types of models traditionally have been developed to represent the events related to use of transportation facilities. These are:

- Land use/land activity models to forecast urban development patterns.
- Trip generation models to forecast the absolute number of trips.
- Trip distribution models to forecast travel patterns.
- Mode-choice models to forecast the allocation of travel between highway and transit and
- Route assignment models to examine probable paths of travel.

These have not been developed as specific formulations applicable to any urban area; rather for each type there has been one or more basic formulation which required the fitting (calibration) of parameters based on existing travel in the specific urban area under study. It has been this requirement which has led to the large data collection costs associated with transportation planning. Chart II illustrates the typical sequence in which the models have been used. Chart III shows the history of the development of the present models.

Land Use Model

It has long been recognized that the demand for travel is a function of both the transportation system and the pattern of land use and land activity. An area which is primarily residential will produce trips from home to work, school, shopping, etc., with a relationship between the density of development (number of people per unit area)

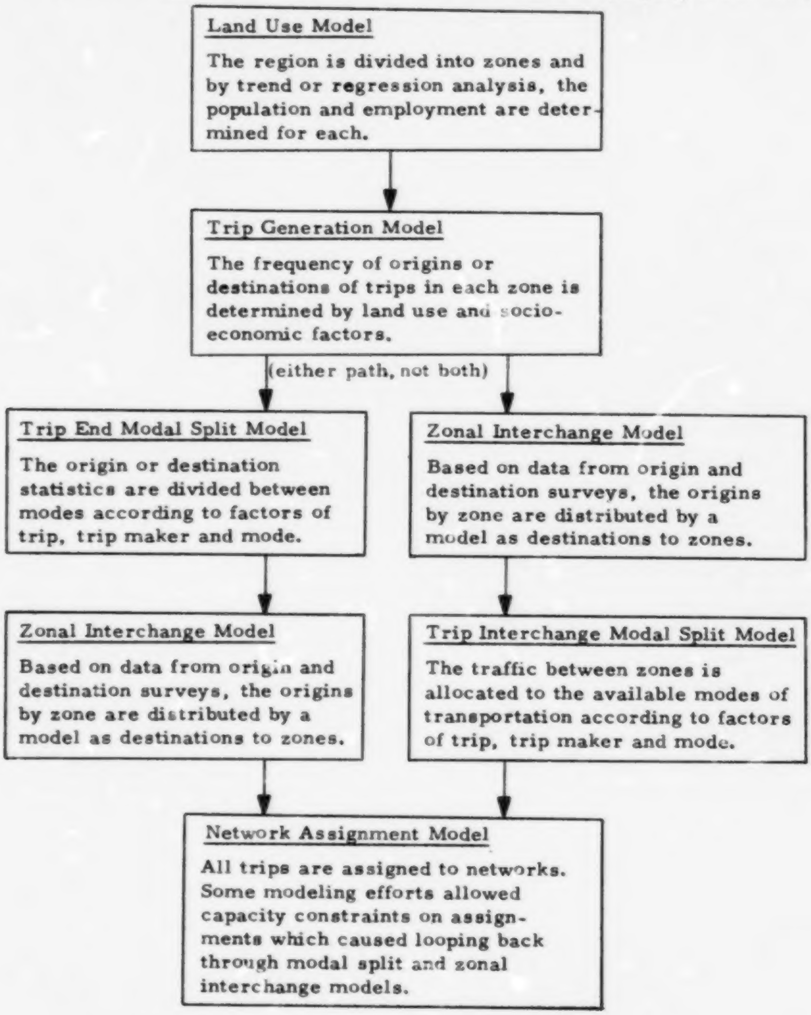

Land Use Model
The region is divided into zones and by trend or regression analysis, the population and employment are determined for each.
Trip Generation Model
The frequency of origins or destinations of trips in each zone is determined by land use and socio-economic factors.
(either path, not both)
Trip End Modal Split Model
The origin or destination statistics are divided between modes according to factors of trip, trip maker and mode.
Zonal Interchange Model
Based on data from origin and destination surveys, the origins by zone are distributed by a model as destinations to zones.
Zonal Interchange Model
Based on data from origin and destination surveys, the origins by zone are distributed by a model as destinations to zones.
Trip Interchange Modal Split Model
The traffic between zones is allocated to the available modes of transportation according to factors of trip, trip maker and mode.
Network Assignment Model
All trips are assigned to networks. Some modeling efforts allowed capacity constraints on assignments which caused looping back through modal split and zonal interchange models.

CHART II—*General Urban Area Transportation Planning Model*

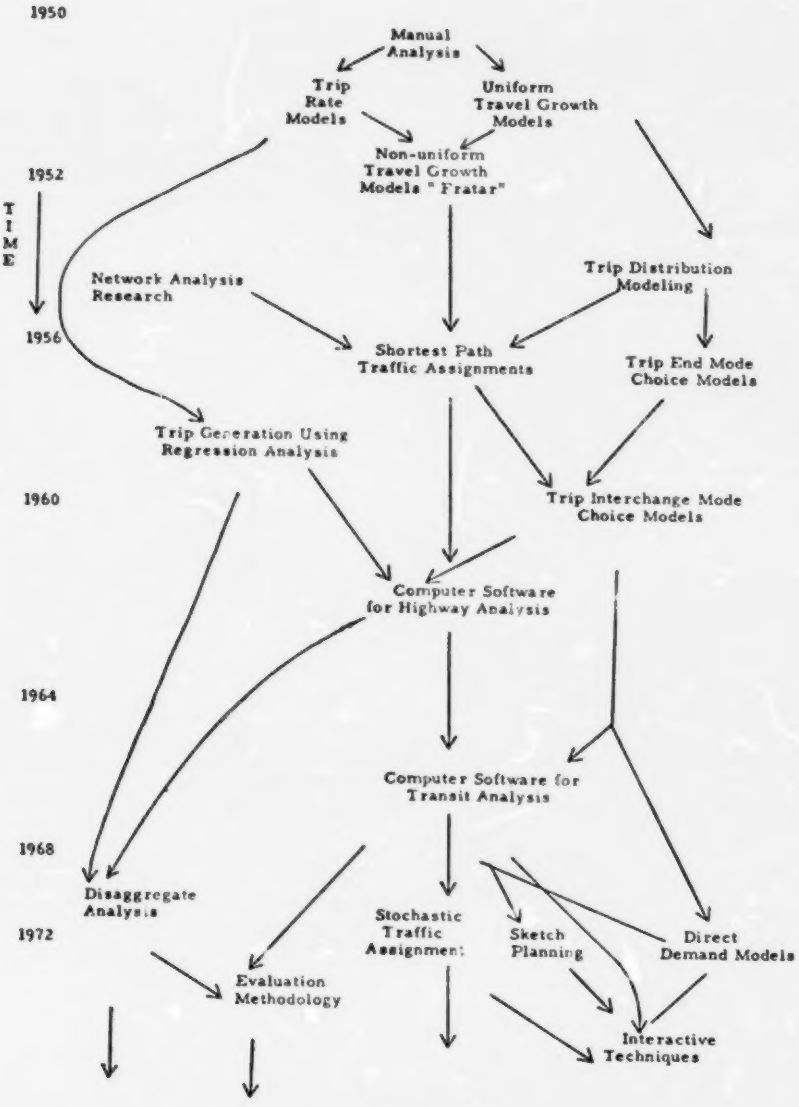

1950
Manual Analysis
Trip Rate Models
Uniform Travel Growth Models
Non-uniform Travel Growth Models "Fratar"
1952
TIME
Network Analysis Research
Trip Distribution Modeling
1956
Shortest Path Traffic Assignments
Trip End Mode Choice Models
Trip Generation Using Regression Analysis
1960
Trip Interchange Mode Choice Models
Computer Software for Highway Analysis
1964
Computer Software for Transit Analysis
1968
Disaggregate Analysis
1972
Stochastic Traffic Assignment
Sketch Planning
Direct Demand Models
Evaluation Methodology
Interactive Techniques

CHART III—*Development of Transportation Models*

and the number of trips. Similarly an employment area will attract trips to work.

On the other hand, it has also been recognized that the development pattern of a region is dependent, in part, on the transport system. Construction of major facilities, such as freeways or rail lines, will open an area for development. Thus there is need for considerable feedback between transportation and land use planning.

One approach to this problem has been the development of models of regional development patterns—land use models—which incorporate as independent variables the existing land use, the proposed transport system, zoning policy, community service infrastructure, and the like. Not all communities have used such models and not all models have required computerized application. Some areas have developed their land use plan on the basis of exogeneously stated goals and then attempted to design a transportation system which would complement the land use. Other areas have modeled the feedback between transport and land use, but with simplistic techniques which can be applied manually.

Bay Area Transportation Land Use Model

In general, a land use model determines for each zone in a region the future population and employment distributions based on given estimates of regional totals for employment and population. In addition, housing, schools, and retail services may be forecast.

In order to better understand the workings of land use transportation models, two of them will be described. The first is a model developed for the Bay Area Transportation Study Commission (BATSC) [15] which was concerned with the development of a regional ground transportation plan for the nine county area around San Francisco as a function of the location of basic industrial employment, location of population-serving employment and location of households. Eight industry types are allocated to zones based on regression analysis of growth trends, accessibility, slope of land, elevation, water frontage, rail lines, employment density, land use today, and current share of county employment. The coefficients of regression based on current and past data are used for the 1990 period.

The allocation of population-serving employment and households to zones is based on estimates of probabilities of trips to other zones using general formulas for trip purposes. The trips are of three types: home to work, home to shopping and work to shopping.

Using this information and industrial location, the model allocates population to zones based on employment in certain industries. The allocation of population-serving population is a function of the allocation of industrial population and the related households. The modeling is dependent on linear regression to estimate parameters, but also make use of deductive relationships such as an assumed form which describes the probability of an individual living less than time t from place of employment.

As noted in the chapter on Urban Models, land use models are quite complex and require an extensive amount of data (that chapter also critiques the efficacy of such models). The major outputs of a land use model—employment, population and households by zones—estimated over the planning horizon, are the inputs into the trip generation models which calculate the number of trips in and out of a zone.

EMPIRIC Land Use Model

Another transportation land use model that uses somewhat different techniques from the Bay Area model is the EMPIRIC model designed for the Washington Council of Governments [30]. The model consists of a set of simultaneous linear equations relating the changes which occur over time in the distribution of population and employment, within a region, to the original distribution of such activity in a given base year, to overall growth in regional activity and to the effects of exogenous planning policy and investment decisions. Activities are defined in terms of district-level measures of population, employment and land use; planning policy measures are expressed in terms of changes in highway and transit accessibilities, utilities, service, zoning and open-space controls.

The linear equations relating the

zone-by-zone shares of total regional population and employment are for a ten year period. They are then used recursively for a total of thirty years. The totals are developed outside of and independently of the land use model. The population is divided into age, income, and size distributions, while employment is divided into ten types.

The methods for developing the model are to use linear regression on data for 1960 and 1968. EMPIRIC is composed of N linear equations in $2N + M$ terms, where N is the number of variables of total value or change of value and M is the number of policy variables. In order to solve this by simultaneous equation solution methods, some factors are reduced to 0 by judgment which produces N equations in N unknowns. The model is designed to yield district-level forecasts of the regional distribution of population and employment for the years 1975, 1985, and 1995, in a form suitable for use as input to the Council of Government comprehensive land-use, transportation, community resources, public safety and health planning activities. Each district/time period forecast includes employment in the areas of industry, retail and consumer service, Federal office, nongovernment office, and others; and population by age, income and family size distributions. Again, we note that these projections are inputs into the next class of models to be described, the trip generation models.

Unlike many other aspects of transportation modeling, no single recommended approach to land use planning has been accepted for use in all areas, nor have Federal agencies funding these studies suggested a recommended approach. Thus, a general land use model formulation cannot be explicitly described although many references are available on several of the larger models.

Trip Generation Models

A trip generation model projects the number of trips made to or from a given area based on measures of activity within the area such as per capita, by household, by acre, by worker, by dollar of retail sales, by square foot or floor space or any other standard unit of land use. In general, it is assumed that the trip rates depend on the type and amount of activity in a zone. Residential, commercial and industrial activities yield different numbers of trips in each zone.

A typical example of a trip generation model was developed for Southeastern Wisconsin [15]. In this model, trip destinations are estimated by regression equations where the independent variables are total employment in retail services, automobiles available, total employment, and retail and service acres.

There is a different linear expression for each of the trip purposes of home to work, home to shopping, home to other, and nonhome based. In most cases these equations have treated each subarea (zone) as a data point. More recent developments have been in the area of disaggregate analysis which treats each dwelling unit as a data point.

Chart IV outlines the independent and dependent variables that are used in the Southeastern Wisconsin trip generation model. The source for historic trip description data is the highway survey where the frequency, purpose, trip origin and trip destination statistics are collected. These surveys are a large component of the overall costs of transportation models.

Another technique for trip generation developed in the early 1960's by the Puget Sound Transportation Study is called "cross-classification" analysis. Rather than using regression to determine linear relationships, this technique involves classifying each zone or dwelling unit according to its trip making related characteristics such as persons per dwelling unit or employees per dwelling unit. This may be a multidimensional stratification. The mean trip rate is then computed for each classification cell. In application zones or dwelling units are multiplied by the appropriate cell mean trip rate to ob-

A_{jp} = transit attractions in zone j of trip purpose $p = 1,2,3,4$

$p = 1$—home based work
$p = 2$—home based shop
$p = 3$—home based other
$p = 4$—nonhome based

$$\boxed{A_{j1} = a_{11} + a_{12}(SE) - a_{13}(A) + a_{14}(E)}$$

where (SE) is total employment on retail service land
(A) is automobiles available
(E) total employment

$$\boxed{A_{j2} = a_{21} + a_{22}(RE)}$$

where (RE) is retail employment on retail & service land

$$\boxed{A_{j3} = a_{31} + a_{32}(SE) - a_{33}(SA) + a_{34}(E) + a_{35}(RE)}$$

where (SA) is retail and services acres

$$\boxed{A_{j4} = a_{41} + a_{42}(SE) - a_{43}(A)}$$

(These are solved by regression on historic data)

CHART IV—*Trip Generation Regression Equations*

tain a total trip estimate. The advantage of this technique is the ability for analysis at a disaggregate level and possibly for development and use of non-linear relationships.

Zonal Interchange Model

This model estimates trips between subareas from input of trip endings, either origins or destinations, in each subarea. The input to the model for each zone is the number of origins or destinations obtained by surveys or trip generation analysis. The earliest trip distribution methods simply applied constant trip rates to the survey results. An improvement in this technique was the Fratar model, which allowed each zone to have its own rate of trip generation. This has evolved into a model calld the "gravity model" which based trip transfers on concepts of attractiveness of a zone for trip ends and impedance between the originating and destination zones.

This model derives its name from its formulation which is similar to Newton's gravitational equations. In brief, the model states that the number of trips between any two areas is directly proportional to the amount of travel activity in each and inversely proportional to the difficulty of travel between the areas. In actual practice this basic concept is modified to reflect social characteristics, travel paths by alternative modes, etc.

A typical gravity model is illustrated by the following equation [16]:

$$T_{ij} = \frac{T_i A_j F(D_{ij}) K_{ij}}{\sum_{j=1}^{j=n} A_j F(D_{ij}) K_{ij}}$$

where the factors of the equation are:

T_{ij} = trips produced in zone i and attracted to zone j.

T_i = trips produced in zone i.

A_j = trips attracted to zone j. frequently a function of floor space or acres of land.

$F(D_{ij})$ = travel time or "friction factor" expressing spatial separation between the zones. This is a monotonically decreasing function of travel cost.

K_{ij} = an adjustment (or calibration) factor to incorporate social, economic, or other factors. There can be more than one adjustment factor.

In general, this formula is applied for each (i, j) combination in an empirical manner, adjusting factors to get reasonable results.

A second type model that is in current use is the "intervening opportunity model." This takes the following form:

$$T_{ij} = O_i \, [P(V_{j-1}) - P(V_j)]$$

where,

T_{ij} = trips from zone i to zone j.

O_i = number of trip origins at zone i.

$P(V)$ = the probability that a trip will end by the time V possible destinations are considered.

V_j = the trip destinations reached before reaching zone j where all zones are ordered in order of increasing travel time from i.

This formulation allows less probability of ending in zone j as the number of intervening opportunities increases.

Trips are made by citizens to suit their needs of the moment. They may be in a hurry, desire comfort or privacy, or may desire many other aspects of travel of a more subjective nature. Often the traveler has a choice of modes to suit his desires, such as rapid transit, private car or a public bus. In the planning process, one must determine the allocation of trips from zone i to zone j among the available modes considered in the planning exercise. The models that determine the allocation are called modal split models [31]. Modal split models are generally either "trip end" or "trip interchange." We next discuss each in turn.

Trip End Modal Split Models

Trip end models allocate portions of either trip origins or destinations to alternate modes of travel based on characteristics of the origin or destination. Trip end modal split models were first designed for highway studies. The Chicago Area Transportation Study [17] first used this technique in 1955. The method was to project subway trips to 1980 based on 1956 rates, then project bus transit trips and finally project auto trips from zone i as a residual. Input to this model is composed of four factors; subway rates, population projections, auto occupancy projections and quantity of autos per family. It is based on trip origins. It should be noted that the modal split model is made up of three submodels, the central transit submodel, the local transit submodel and the automobile traffic submodel. Central transit is defined as trips ending in the Central Business District. Local transit is all other.

Trip end modal split models have become more complicated with the use of increased numbers of factors. An example of a more complicated and more sensitive and accurate model is a trip end modal split model called the Southeastern Wisconsin Regional Planning Model [15]. Tripmakers are categorized into four kinds and trip attractions are calculated by linear regression on factors such as employment, automobiles, retail and service acres. Trip characteristics are calculated based on friction factors, a function of door-to-door travel time. Transportation systems characteristics are differences in trip times. This model uses the outputs of the previously described Southeastern Wisconsin trip generation model as inputs.

Trip Interchange Modal Split Models

These models allocate the portions of trip movement resulting from trip distribution to the alternative transportation modes. The earliest and the most simple is a procedure designed for Washington, D. C. in 1960 [31]. The concept is that the input to the model, the total number of people moving from zone i to zone j, are a market that the modes compete for. In the Washington model, the factors of travel time cost, economic status of the trip maker and relative travel service are used for evaluation. Trip interchange modal split models in other cities have used other variables such as employment density, workers per household, residential density, auto ownership, length of trip, and time of day. Essentially, in each case, the estimate of the proportion of transit trips from zone i to zone j is either based on linear regression or nonlinear table look up. Both estimation methods depend on a great deal of historical information.

The detailed data requirements for such models are illustrated by the Washington, D. C. models need for the following: the total interzonal trips for work and nonwork, travel times between zones by auto and public transit by purpose of trip, transit fare, waiting times for transit and parking an auto, auto parking costs, income levels of travelers, average auto occupancy, auto trip costs between zones and by purpose, and walking time for transit and auto. An interesting feature of this model was the use of diversion curves which calculated, for example, the percentage of trips made by transit as a function of travel time.

Network Assignment Models

The final step in the process of estimating trips for a planning horizon is to assign the trips to the road and transit networks. In the early transportation planning efforts the assignment to the network of a trip from zone i to a zone j was done by judgmental decisions as to the "most likely" route. However, techniques were developed in 1956 for finding the minimum cost or minimum time path through a network [62]. On the assumption that people are all desiring the same thing, to make minimal time or minimal cost routes, these approaches are now an accepted method of assignment of trips to a network. It has also been used in an iterative manner by altering trip times on links so that all traffic does not get put on the highest capacity links, or using "diversion curves" which apportion trips among shortest time and shortest distance paths.

A recent advance in route assignment eliminates the concept of a single path between two zones and assigns trips on a probabilistic basis to all routes within an envelope of feasible paths.

Deficiencies of the Traditional Approach

The transportation modeling process as it has developed over the past twenty-five years has served a purpose in that it has laid a groundwork for systematic analysis of urban problems and has, to an extent, assisted policy makers in making rational decisions. Unfavorable public reaction which now attends many transportation decisions indicates that there are deficiencies both in the modeling process and in the way in which model results are interpreted.

At the base of these deficiencies has been the inability of planners and decision makers to clearly formulate a set of regional and subarea goals which are compatible. Aside from general platitudes such as maximum opportunity or freedom of mobility the planning profession has not been able to formulate measures of desirable development patterns which meet general public approval. Thus, transportation plans have been structured, for the most part, to continue observed trends in community development and lifestyle. This basic formulation of regional growth is reflected in the model results in the form of ever increasing demands for new highway facilities. When these model forecasts are evaluated in regional terms without consideration of attendant subarea costs the results almost without question indicate the need and desirability of massive highway programs. While many highway improvements are clearly called for the results of the modeling process has been to overstate the need and to avoid consideration of alternatives which cannot be readily simulated. The result has been public dissatisfaction with transportation plans and transportation planners.

More specific problems arise from the basic model structure. The traditional chain of models treats each travel decision—to make a trip; the destination chosen; the mode used and the route used—as an independent decision. In actual fact a tripmaker considers all facets simultaneously.

Trip generation models typically are not sensitive to the supply of transport services. According to the models building new facilities such as a freeway or rapid transit line will not, of itself, result in more trips by the existing population. In fact, increased accessibility does result in more trips.

Trip distribution models have several failings primarily due to the level of aggregation. In few cases has the forecast of travel patterns considered the effect of transit in choice of a trip destination. Only a few attempts have been made to consider the characteristics of individual trip makers.

Choice of mode—auto driver, auto passenger, transit passenger—has been the subject of much research over the past decade. Models of mode choice now can yield reasonable estimate of usage for traditional transit systems, but give little guidance on the potential for the many new concept transit proposals now under consideration. Similarly the models, requiring extensive data basis and costly operation, provide little guidance in transit system design.

A major failing has been in models for system evaluation. Considerations have been treated at a regional scale with major attention on direct costs and benefits—those items which are simulated by the models. Little attention has been given to economic and social aspects of transport facilities, the impact of neighborhoods disturbed by facilities or the consequences of development resulting from the facilities and the attendant demands on other aspects of community intrastructure.

Finally, the transport models have not been constructed to consider equilibrium situations. In fact, there is significant feedback among all aspects of the problem. Land use and transport are intimately related. Highway congestion impacts mode choice. Increased parking demand is reflected in increased travel cost. The cost and time involved in applying the complex models, however, effectively precludes both cycling of model application to equilibrium and consideration of many alternative systems. It is perhaps this latter failing—the inability to examine alternatives within reasonable times and costs—which has had the most significant impact on the planning process.

Estimates based on the experience of large scale urban studies indicate that approximately 50 percent of the cost is developed to data collection and data processing with only 15 percent devoted to evaluation. This is due to the massive data requirements of the current models. With this model structure planners have found that the cost and time required to effectively treat simple problems preclude analysis in a reasonable time frame. Thus, planners have not been able to provide timely response to the public or to political bodies such as city councils or the Congress. For truely effective planning, it is imperative that response time and cost be reduced so that analysis may be provided to these public forums.

Recent Developments and New Directions

Since the late 1960's most of the deficiencies discussed above have been recognized by planning professionals and research has been underway to develop techniques which consider the untreated aspects of planning, are more responsive to the political decision making process and may be applied more quickly and cheaply. We discuss a few of these next.

Multi-Modal Planning

Among the techniques now being actively pursued are those which will facilitate more accurate treatment of systems containing non-highway elements. These multi-modal models permit more effective treatment of both traditional and new concept transportation facilities minimizing the highway bias which has been engendered by the availability of sophisticated highway planning models without comparable transit planning models. This new emphasis is illustrated by the UMTA Transportation Planning Package now available to and in use by local planning agencies providing a framework for comprehensive transportation analysis [73, 74].

Demand Estimates

The lack of interrelationship between the several elements of the traditional model chain is being attacked through the development of "direct demand models" [53]. This class of models, which are still under development, attempt to simulate simultaneously all aspects of the travel decision including generation, distribution, mode choice, and even route selection. In particular, these models attempt to overcome a major technical deficiency by relating the demand for travel to the supply of transport services.

In addition the direct demand models, as well as the newer versions of the traditional models, place more emphasis on disaggregate analysis. Rather than treating large subgroups as a uniform mass the disaggregate models recognize that each individual perceives transport choices differently and, thus, makes different decisions. Only with such an approach, it is believed, can models accurately simulate the totality of transport demands.

User-Oriented Planning Techniques

Other groups are devoting attention to techniques which will permit cheaper and more rapid consideration of alternatives so that technical data can be produced in a time scale commensurate with public and political demands. These efforts are focused on increasing speed through use of less complex analysis (sketch planning) [56] and more effective utilization of the existing techniques and large scale digital computers (interactive planning) [58, 63, 68].

In sketch planning the premise is that few major transport investment decisions require absolute detail and that, in fact, too much detail may obscure the primary issues. By increasing the scale of the problem the planner requires less data and faces a less complex situation so that many major transport corridor alternatives and types of transport facilities can be screened quickly and cheaply leaving only a few feasible alternatives requiring detailed analysis.

Interactive planning reduces the time scale for simulation so that in a few hours a planner may evaluate a host of alternatives keeping the results available for rapid evaluation and modification. Thus the planner can profit from experience to build upon prior findings in developing new alternatives. This technique offers great promise in effective utilization of human resources for planning as opposed to mere data manipulation.

Data Sources

To reduce costs of planning increasing emphasis is being placed on the use of available data resources. Millions of dollars have been spent in urban area data collection over the past decade. This need not be repeated on a periodic basis. Intragovernmental cooperation of the use of census data and other on-going programs combined with statistically well designed subsampling will enable models to be developed or validated at considerably less cost.

Other Areas

Much attention is now being given to new techniques to deal with urban travel problems. These include better use of existing street systems through demand responsive traffic control system bus priority lanes and the like. Models to deal with this class of problem are now under development. Other techniques focus on the use of new concept, new technology concepts such as dual-mode vehicles, dial-a-bus, personalized rapid transit small vehicle systems, paratransit, etc. To date no adequate methodologies have been developed to model these systems.

III. INTER-URBAN TRANSPORTATION PLANNING

One of the most important models to be developed in the field of transportation was the Northeast Corridor Transportation model. The northeast megalopolis is dependent on transportation for movement of freight and passengers. The megalopolis could not exist without an efficient system of traffic flow. It was recognized by the Federal government in the early 1960's that improvements were necessary because of the ill health of the railroads and the congestion of the highways. MIT was contracted to propose new innovative technology such as tube transport, high speed surface transport and other new techniques. It was found necessary to review investment planning and public policy to determine the effects on the corridor transportation. It was determined that it was imperative that assessment be made of the future environment and the usefulness of new systems and new policy in the environment. A model was necessary for decision making on the following type questions:

- The effectiveness to the Northeast Corridor of proposed new and advanced technology, such as the Metroliner service now in operation
- To determine the operational characteristics and economic viability of future transportation technology
- To determine the changes in transportation systems that will be necessary to support projected shifts in industry and population
- To determine the facilities that will be necessary to support projected shifts in inter-modal handling of freight and passengers
- To test the effects of public policies on passenger and freight systems, such as the development of AMTRAK

The complexities of these decision problems led the Department of Commerce to consider modeling as the method of arriving at answers. The Northeast Corridor Model was the result [27], [28].

The objectives of the project were to develop a framework of analysis of transportation systems for general use in transportation planning and to reduce the uncertainty in planning of public investment and transportation policy in the Northeast Corridor Region. The Region consists of those states which have highly urbanized areas connected or nearly connected to one another along the eastern seaboard of the United States. They are New Hampshire, Massachusetts, Rhode Island, Connecticut, New York, New Jersey, Pennsylvania, Delaware, Maryland, District of Columbia and Virginia. This is roughly the Northeast megalopolis.

The states of the Northeast Corridor are some of the most highly urbanized in the United States. The majority of the population is concentrated in the cities which are manufacturing and trade centers such as Boston and New York.

In the description of the Corridor model which follows, the overall model is made up of nine submodels, as shown in Chart V. The workings of the submodels will be described to illustrate the complexities of a model as vast as this.

The Corridor Model is designed to generate forecasts of the levels of economic activity in the States. These forecasts comprise "control totals" for models which generate the forecasts of allocation of economic activity and population to a lower level of aggregation. The flow from the Corridor Model to the Inter-Regional Input-Output Model consists of gross product by ten industrial sectors and the estimates of consumption. The flow also includes private non-

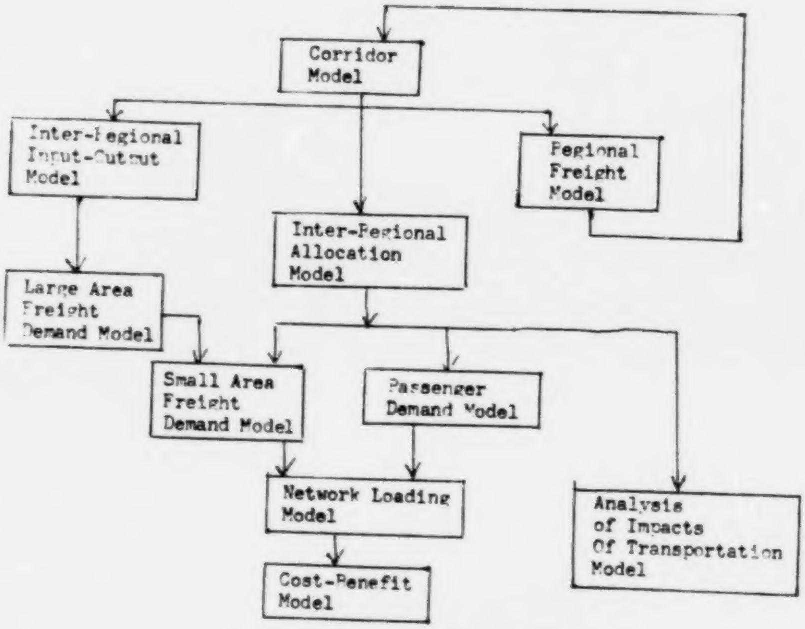

Corridor Model
Inter-Regional Input-Output Model
Regional Freight Model
Inter-Regional Allocation Model
Large Area Freight Demand Model
Small Area Freight Demand Model
Passenger Demand Model
Network Loading Model
Analysis of Impacts Of Transportation Model
Cost-Benefit Model

CHART V—*Northeast Corridor Model*

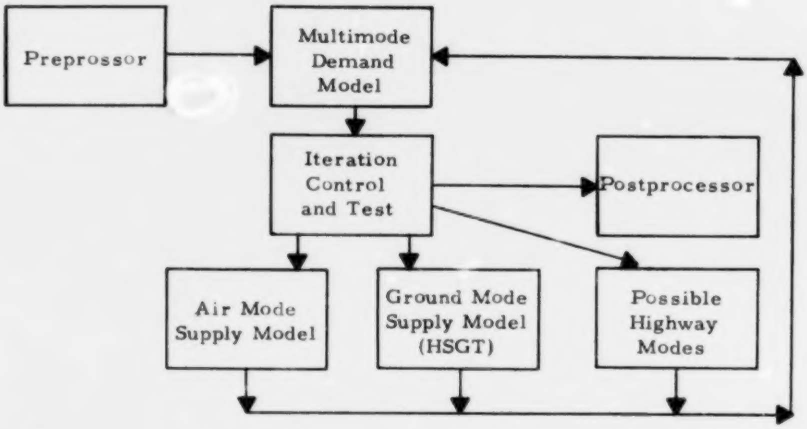
farm residential construction, gross fixed nonresidential investment, purchases of goods and services of State, local and Federal governments. The forecasts of gross product by industry comprise the information necessary to establish the product for six subregions within the Corridor and the region outside of the Corridor, but within the States of the Corridor Project. The expenditure flows provide the information necessary for the establishment of the final demands for the regions.

The Inter-Regional Input-Output Model is an operational version of the model constructed by Wassily Leontief and Alan Stout [29]. It disaggregates the totals for each of the three regions, north, central and south into six smaller regions with the Corridor, by the use of technological coefficients and parameters which characterize interregional commodity flows, reflecting demand and the quality of the transportation network. The outputs from the Inter-Regional Input-Output Model flow into two other models. First, a model to allocate population, employment and income to approximately 100 subareas of the Corridor. Second, a model which generates the demand for freight transportation services on the basis of the predicted inter-regional commodity flow.

The Inter-Regional Allocation Model allocates population, employment and income on the basis of a system of equations which take explicit account of the influence of changes in the overall transportation network. It thus provides the basis for forecasts of the spatial pattern of regional economic development as it is influenced by transportation. Employment is predicted by industrial sector. Personal income, wage differences, and the employment compositions are predicted for the 100 subareas.

The forecasts from the Inter-Regional Allocation Model are variables in the model which predict passenger travel demand. Passenger travel demand between two areas is to a large extent determined by the levels of economic activity in the two areas, as well as the characteristics of the transportation network. The Inter-Regional Allocational forecasts are used in the prediction of freight demand among the 100 subareas by using characteristics of the transportation network.

The models as described were never completed in the forms described but were supplemented by approximation methods and by other models. One model in particular was of great supplemental use, the Multi-Modal Transportation Model System [77]. This model is described in the following flow chart.



When the demands for transportation services between origin and destination pairs have been predicted, they are allocated to routes within the network using a computer simulation. Information on predicted transportation user costs per transportation unit and predicted volumes of transportation using the network routes are converted into measures of transportation users' costs and benefits. Costs of the operators of the transportation system are estimates, and a cost benefit analysis of the direct effects of alternative transportation systems is performed.

The complexities of each of the submodels of the Northeast Corridor project may not be apparent in the brief descriptions just made. The input-output scenarios which are the basis for testing alternative transportation systems are composed of many variable estimates. The ranges of error of estimate for the cost benefit analysis results are difficult to predict or even to roughly estimate.

This was an iterative model where the demand model initiated the estimation and supplied the mode operators. These operators made up service levels without knowledge of what the competitors are doing. The operators observe the results from the last iteration of the results of their own and their competitors' actions. This creates a new demand and a new iteration. This cycling continues until the change from cycle to cycle becomes negligible. The service characteristics of the modes were altered in successive runs to allow the modes a reasonable return on investment and profit maximization.

IV. NATIONAL TRANSPORTATION PLANNING

The Department of Transportation planning for transportation on a national scale is generally in two decision making areas.

- Under existing legislation, to determine the disbursements of monies to State and local governments which best meet the investment needs and program priorities. The Federal government is committed to provide increased flexibility in Federal Aid programs, as contained in the Federal Aid Highway and Mass Transportation Act of 1972.

- To determine new legislation to propose to Congress to modify the Federal role as planning uncovers new and special problems.

The 1972 National Transportation Report [11] is the first in a series of reports of the state of transportation in the Nation, and the plans of Federal, State and local governments for improving the systems. As noted in the 1972 Report, since 1957, 71% of the Federal monies have been spent on highways. Recent increases in use of Federal monies has been noted in urban mass transportation and aviation. The Federal government has been concerned mostly with developing highways, but now is getting increasingly involved in local transportation. Only recently has the Federal Government spent anything on the railroads. In addition, the Federal government is concerned in its comprehensive planning process with the depletion of the energy reserves of the nation. Transportation consumes one fourth of the energy each year. Safety, air hijacking and cargo security are also major concerns.

The States and local planning groups were asked to present their transportation needs and the results are included in the 1972 Report. In the aggregate, the States indicate a need for $670 billion for their needs of which 84% is for highways. These requirements are not constrained by the reality of funding availability, energy availability, environmental degradation of alternative land use priorities.

In order for the DOT to assess the nation's transportation needs, several models were used in preparing the 1972 Report, including the Airport Investment Model [10] and a multi-mode national transportation model [12].

In addition to the National Transportation Report, the DOT is required to provide a National Highway Needs Report. This involves a field inventory in each State of a representative sampling of highways. These reports were published in 1968 and 1970. The National Transportation Report went beyond this and did the following:

- Asked each State to develop and report alternative capital improvement programs subject to different realistic Federal fund levels and degrees of flexibility.
- Forecast travel and freight flow by mode based on anticipated growth rates of the economy, special requirements such as defense and the relationship between the economy and traffic flow.
- Surveyed most transportation facilities and services largely by State submissions, and performed economic analyses to determine the types of improvements likely to be most profitable or of most value to society.
- Surveyed the private sector and other governmental agencies to obtain estimates of future capital investment requirements in these sectors.

Another required planning effort results in the National Airport Systems Plan and is directed at a ten year plan for the development of the public airports in the United States, to meet the growth of aviation. These three reports reflecting DOT planning efforts are required by existing legislation.

A Transportation Assessment Model

A planning model, TRANS [8], was developed in response to the evaluation needs for the National Transportation Report and for assessing the disbursements of Federal monies to the States. TRANS is a national model which requires as input the result from a massive national survey. This survey is composed of 244,000 samplings of highways, arterials, connectors and local roads. Seventy-four parameters are collected on each sampling, including accident rates, pollution produced, average speeds, amount of traffic, travel time, design characteristics, fatalities, noise, aesthetic problems, etc. Standards are given to determine if a sampling passes or fails on each of the parameters. The model performs cost benefit analyses on possible improvements, measured in terms such as lives saved or lost, and impact on the economy. The cost benefit ratios are then ranked by State to gain a first approximation of the best use of Federal funds. It also provides a method of evaluating the proposals that come from each state for extensions and improvements.

The TRANS model is an aggregative systems. In the urban areas of over 50,000 population analysis is done by individual highway units. Smaller urban areas from 5,000 to 50,000 population are grouped by State, while rural areas are analyzed by the aggregate for the whole State. A simplified flow diagram for the TRANS model is shown in Chart VI.

TRANS is really a large computer analysis system following the steps in Chart VI. The statistics on the 244,000 highway samples of 74 parameters are fed into the subprogram that determines the alternative remedies where systems have failed standards. Based on related socioeconomic data and the selected remedial alternative for the highway, travel demand is projected. The alternative highway is costed with regard to renovation, new construction or extension. The cost savings to users is determined by mileage savings. The monies invested in the proposed alternatives are compared to the economic and other benefits. Cost benefit ratios are computed for each highway link and they are listed by ranking order

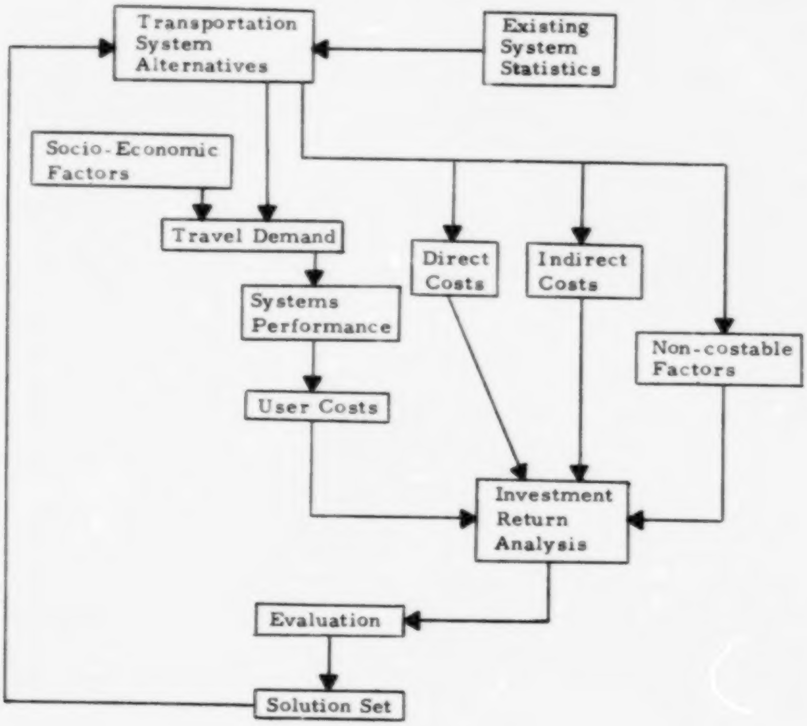

Transportation System Alternatives
Existing System Statistics
Socio-Economic Factors
Travel Demand
Direct Costs
Indirect Costs
Systems Performance
Non-costable Factors
User Costs
Investment Return Analysis
Evaluation
Solution Set

Chart VI—*The TRANS Model*

for each State. The highest ranking cost benefit ratios are the projects which yield the most benefit for the proposed investments.

The model results have not matched the submittals from the States on their highway needs, indicating a lack of conformity between the two levels of government with regard to objectives. The objectives for most local areas are not well defined and in some cases do not exist.

TRANS has been used for policy analysis as well. Some of the policy results have been the following:

- If areas of one million population or more staggered work travel times uniformly over a 3 hour morning and evening time periods, this would bring about a 25% reduction in required freeway miles.
- In urban areas, an additional payment of $5,000 for each dislocated family would reduce freeway miles by 7% under existing funding constraints.
- If the economic value of travel time were to be doubled, it would result in doubling the possible national arterial investment.
- If the cost per fatality were increased by $100,000, urban investment levels would increase by 2% and rural investment levels by 4%.

Airport Investment Model

In support of the planning effort of the National Airport Systems Plan for the development of the airports of the United States, an airport investment model has been developed [10]. Airports exemplify the complex nature of public utility investment and pricing decision making, which ranges across institutional, economic and operational considerations. Federal bureaus regulate the air carrier industry, while State and local commissions supported by Federal advice make the airport capacity decisions. The public demand for service is related to the capacity of the system and the behavior of the carriers. Demand for service is not uniformly distributed over time in that intensive utilization occurs for several hours of the day and the demand is seasonal.

In the study about to be described, capacity and cost allocation decisions with respect to public airports are considered. A model is used to quantify the benefits received and the capacity required by various airport user groups. User group demand elasticities are approximated by hypothesizing their response to the costs and benefits of airport operating strategy alternatives. From this, measures of effectiveness of system improvement policies are obtained. Approximate measures of subsidization of peak users by off-peak users and of general aviation by the air carriers are also obtained. The model and the study results were input to a major transportation study [13] as a basis for new legislation. The model has not been used as yet for individual airport decision problems.

In this model, passenger demand is an input for a particular airport being analyzed. Passenger demand is projected from 1970 to 1990 and in most cases is assumed to be increasing. It is also independent of the facilities offered at the airport and the airline schedules.

The problem addressed by the model is to find the optimum sequence of airport investments in capacity to meet the growing demand. There are two ways for obtaining capacity increases:

- capital expenditures such as some additional runways, more extensive air control, construction of new airports and development of a separate short haul system, and
- non-capital alternatives include diversion of general aviation to general aviation airports and spreading of peak activity into other time periods.

Analysis of the capital expenditures requires an assessment of the total cost of usage of air carrier airports. Analysis of the peak spreading in the non capital alternative requires assessment of demand elasticities as related to the timing with subsequent inconvenience costs.

The model has two major parts:

- the calculation of total cost including capacity cost, air carrier aircraft delay cost, passenger delay cost, general aviation delay cost and general aviation passenger delay cost. Capacity costs are related to the design of the airfield and the ratio of air carrier operations to total operations, and
- a network analysis model of feasible expansion designs on a year to year basis. There are a large number of combinations of possible improvements of airports that can be made on a year to year basis. For instance a new runway can be added in a certain year and a new air control system the following year. Or they can be reversed, neither implemented or only one of them. That results in seven combinations for this small problem. The model objectives is to find a yearly sequence of improvements to the planning horizon that minimizes the total cost.

The network analysis is done with the dynamic programming technique. Dynamic programming is a procedure for selecting a sequence of decisions which yield the best solution over a time span. The optimal decision process in this case is the sequence of investments in capacity to a planning horizon of 1990. Each decision is the determination of the amount of capacity to add to the airport each year. Each year has a number of possible designs for the airport associated with it. The design possibilities for the n^{th} year depend on the design adopted for the $(n-1)^{th}$ year, including no change from the previous year.

An Aircraft Route Effectiveness Model

The Federal government has requirements for models to aid in the evaluation of airline operations in terms of policies of fares and user charges, route awards and requests for service changes, mergers and intercarrier agreements, new terminal decisions, limits on airport operations, effects of new aircraft, and the impact of future growth. The study and modeling effort about to be described were aimed at developing a method to approximate the mix of planes, routes, schedules, and terminal facilities that would satisfy intercity passenger and cargo demand at minimum social and economic costs [9].

The flowchart of the model is shown in Chart VII. The heart of the system is the network assignment module, which assigns passengers, cargo and aircraft to routes to minimize total cost. Total cost is composed of airline cost plus passenger and cargo time cost. The output of this module is the daily frequency of service required on each route, along with the passenger (and cargo) and aircraft assignments, and a summary of the various costs associated with the solution.

The demand module estimates the total daily passenger and cargo demand for each city pair and determines the share to be allocated to the air carrier under study.

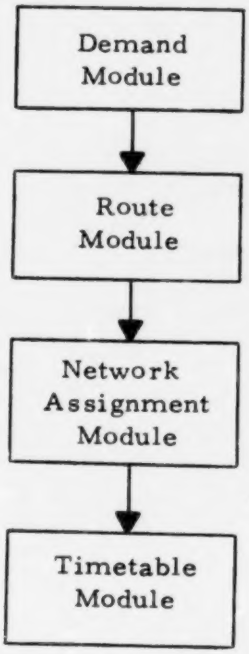


CHART VII—*Aircraft Route Effectiveness Model*

The route module selects the best routes for a subsequent use by the network assignment module. Demand, geography, and route restrictions are all considered.

The timetable module takes the daily frequencies on each route and produces a non-optimal timetable, but one that reduces carrier requirements as much as possible.

The formulation used for the demand module was a gravity model, described earlier. Passenger traffic or demand between two points is hypothesized as directly proportional to the product of the masses of the populations and inversely proportional to the separation between the points which is impedence to travel. The friction factor is an exponential function of:

- terminal to terminal nonstop travel time,
- ratio of actual expected travel time to nonstop travel time,
- access time to and from terminal to local origin or destination (including average terminal processing time),
- frequency of service, and
- fare.

The network assignment model is essentially a decomposed linear program that assigns passengers and aircraft to the routes on the basis of least total cost. The solution consists of the number of flights per day required to satisfy the demand for each origin and destination pair over specific routes.

This model was used in one instance to supply model results for a CAB hearing. The case was a prospective merger of American and Western Airlines. The model results supported the merger in terms of lower total costs and the merger was approved.

Summary

In summary, the use of models in making decisions on integrated national transportation problems is in a beginning phase. The models presented above are designed for particular subsystems of the overall transportation system. In reality the subsystems are in competition with each other for both freight and passengers. The existing laws, rate structures, taxes and subsidation appear to benefit one subsystem more than another. Some subsystems such as road networks are more constrained by local legislation and planning controls than are others. The objectives and priorities in local government are diverse or non-existent, and in many instances inconsistent with national objectives.

These problems point to the need for increased modeling effort on analyzing the entire transportation system, including railroads, shipping lines, buses, trucks, passenger automobiles, aircraft, etc. The required models for such an analytical effort should allow competition among the modes. With an inter-mode competition model, policies and fundings for a given mode can be tested to see the effects on all modes. The concept of a balanced transportation system of all modes could then be studied for the purpose of changing existing legislation and developing consistent objectives and priorities applicable at all levels of government.

References

[1] Federal Highway Administration, "Economic and Social Effects of Highways," DOT, 1972.

[2] Massachusetts Institue of Technology, "Economic Impact of Massachusetts Route 128, Boston, December 31, 1958.

[3] Wilbur Smith and Associates, "Maryland Capital Beltway Impact Study," Columbia, South Carolina, June 1968.

[4] Sawhill, Roy, "Freeways and Residential Neighborhoods—A Research Case Study Seattle, Washington," Automotive Safety Foundation, Washington, D. C., July 1965.

[5] Hill, Stuart L., *The Effect of Freeways on Neighborhoods*, California Division of Highways, Sacramento, California, June 1967.

[6] "Beauty for America," *Proceedings of the White House Conference On Natural Beauty*, Washington, D. C., May 24–25, 1965, U.S. Government Printing Office.

[7] Highway Research Board, *Press Kit*, 52nd Annual Meeting, January 22–26, 1973.

[8] *TRANS*, a not for release model description, DOT.

[9] "Intercity Transportation Effectiveness Model," Peat, Marwick Mitchell and and Company, December 1970, Teprot TB 542–1.

[10] "Airport Investment Model," unpublished document, DOT.

[11] "The 1972 National Transportation Report," DOT, July 1972.

[12] Kassoff, Harold, "A Multi-Modal National Transportation Planning," Transportation Resource Allocation Study, FHA, Washington, D. C., March 1972.

[13] "The 1972 National Transportation Needs Study," DOT, 1972.

[14] Swerdloff, Carl N., "Developing A National Network Model of Intercity Freight Movement In the United States," a DOT paper.

[15] National Bureau of Economic Research, Inc., "Overview of Land Use Transportation Models in Planning and Survey of Land Use Models in Five Metropolitan Areas.

[16] Hansen, W. E., "Evaluation of Gravity Model Trip Distribution Procedures," *Highway Research Board Bulletin No. 347*.

[17] State of Illinois, *Chicago Area Transportation Study Vol. II Data Projections*, July 1960.

[18] Webb, Kenneth W., "Investigation of Human Participation in Northeast Corridor Modeling," December 1965.

[19] Boyce, David E. and Roger W. Cote, "Verification of Land Use Forecasting Models, Procedures and Data Requirements," "Land Use Forecasting Concepts," Highway Research Record, Report No. 126.

[20] Webb, Kenneth W., "The Mathematical Theory of Sensitivity,' A paper presented at the Eighteenth National Meeting of the Operations Research Society of America, October 1960.

[21] Alonso, William, "The Quality of Data and the Choice of Design of Predictive Models," "Urban Development Models", Highway Research Board Special Report 97.

[22] Kain, John F., "A Re-Appraisal of Metropolitan Transport Planning," *The Logistical Review*, Vol. 5, Jan.-Feb. 1969, No. 21.

[23] Harris, Britton, "The Uses of Theory in the Simulation of Urban Phenomena," "Land Use Forecasting Concepts," Highway Research Record #126.

[24] Schneider, Morton, "Access and Land Development," "Urban Development Models," Highway Research Board, Report 97.

[25] Lowry, Ira S., "A Model of Metropolis," RAND Memorandum RM-2035-RC, August 1964.

[26] Whol, Martin, "Another View of Transport System Analysis," Proceedings of the Institute of Electrical and Electronics Engineers, Vol. 56, No. 4; April 1968.

[27] Department of Commerce, Study Design, Northeast Corridor Transportation Project, Technical Paper No. 5, 1966.

[28] Crow, R. T., "An Econometric Model of the Northeast Corridor of the United States," MATHEMATICA, March 1969.

[29] Leontief, W. and A. Strout, "Multiregional Input-Output Analysis," Structural Interdependence and Economic Development, London, MacMillan, 1963.

[30] Peat, Marwich, Livingston and Co., "Empiric Activity-Allocation Model Study Design," Metropolitan Council of Governments, seven technical memoranda, April 1969.

[31] Department of Commerce, "Modal Split, Documentation of Nine Methods for Estimating Transit Usage," Bureau of Public Roads, Washington, D. C., U. S. Government Printing Office, Dec. 1966.

[32] Mitre, "Form and Behavior of a Transportation Demand Forecasting Model," MTP-370, April 1972.

[33] Koike, Hirotaka, "Planning Urban Transportation Systems, A Model for Generating Socially Desirable Transportation Network Configuration," Report #2, University of Washington, Seattle, 1970.

[34] Federal Highway Administration, "Estimating Auto Occupancy, A Review of Methodology," 1972.

[35] Elias, S. E., *A Mathematical Model for Optimizing the Assignment of Man Machine in Public Transit "Run Cutting,"* West Virginia University, September 1966.

[36] Elias, S. E., *The Use of Digital Computers in the Economic Scheduling for Both Man and Machine in Public Transportation*, Kansas State University Bulletin #49, undated.

[37] Landi, D. M. and A. J. Rolfe, *A Model and Computer Code for Studying Alternative Air Passenger Processing Strategies*, RAND RM 5818, August 1969.

[38] "Chicago Area Transportation Study," (Final Report) Vol. I, December 1969, Vol. II, July 1960.

[39] Roos, Daniel, et. al., "Summary Report, The Dial-a-Ride Transportation System," MIT Urban Systems Laboratory, March 1971.

[40] *Studies in Travel Demand*, Vol. IV., Northeast Corridor Project, DOT, May 1969.

[41] Roberts, Kenneth, "Vehicle Scheduling and Driver Run Cutting," Mitre, M71–58, September 1971.

[42] *Urban Transportation Models, OKI* Regional Transportation and Development Plan, December 1968.

[43] Hamilton, William F., "Systems Analysis of Urban Transportation," *Scientific American*, Vol. 221, No. 1, 1969.

[44] Morlok, E. K., "Methodologies in the Selection and Evaluation of Transportation Alternatives," *High Speed Ground Transportation*, Carnegie-Mellon University, 1969.

[45] Boyce, David E. and Norman D. Day, *Metropolitan Plan Evaluation Methodology*, Institute for Environmental Studies, University of Pennsylvania, Philadelphia, Pennsylvania, 1969.

[46] Burco, Robert A., et. al., "Future Urban Transportation Systems: Impacts on Urban Life and Form-Study in New Systems of Urban Transportation," Vol. IV, Stanford Research Institute, Menlo Park, California, 1968.

[47] *Detroit Metropiltan Area Study, Part II: Future Traffic and a Long Range Expressway Plan*, 1956.

[48] Hoel, L. A., "Evaluating Altenative Strategies for Central-City Distribution," *Highway Research Record* 293, 1969.

[49] Manheim, Marvin L., "Search and Choice in Transport Systems Analysis," *Highway Research Record* 293, 1969.

[50] Department of Housing and Urban Development, "Tomorrow's Transportation, New Systems in the Urban Future," Urban Transportation Administration, Washington, D. C., 1968.

[51] Barton-Aschman Associates, Inc. and Peat, Marwick, Mitchell and Company, New Systems Requirements Analysis Program, UMTA, "Functional Specifications for a Transit Station Simulation Model," November 3, 1972.

[52] Brand, D. et. al., "Notes on a General Framework for Transportation Analyses," *Proceedings of the ORSA-TSF Urban Transportation Workshop*, August 1971, pp. 111–119.

[53] Brand, D., "Transportation Demand Forecasting, Some Foundations and a Review," *Conference on Urban Travel Demand Forecasting (UTDF Conference)*, Williamsburg, Virginia, December 1972.

[54] Dial, R. B., "A Probabilistic Multipath Traffic Assignment Model Which Obviates Path Enumeration," *Transportation Research*, Vol. 5, 1971, pp. 83–111.

[55] Dial, R. B., "Demand Forecasting for New Options and Technology," *UTDF Conference*, Williamsburg, Virginia, 1972.

[56] Dial, R. B., et. al., "A Procedure for Long Range Transportation (Sketch) Planning," *International Conference on Transportation Research*, Burges, Belgium, June 1973.

[57] Federal Highway Administration, "Urban Transportation Planning—General Information," March 1972.

[58] Gur, Yehuda, "A Computer System for Interactive Analyses and Planning of Urban and Transportation Systems," Technical Report Number 1, Department of Systems Engineering, University of Illinois, Chicago Circle, May 1972.

[59] Hitchcock, F. L., "The Distribution of a Product from Several Sources to Numerous Localities," *J. Math. Phys.*, 20, 224–230 (1941).

[60] Loubal, Peter S., and Bernard Mendes, "On-Line Travel Demand Forecasting Model," *Proceedings, Urban Traffic Model Research*, P.T.R.C., London, May 1972.

[61] Manheim, M. L., "Practical Implications of Some Fundamental Properties of Travel Demand Models," paper prepared for the Annual Highway Research Board Meeting, January 1972.

[62] Moore, E. F., "The Shortest Path Through a Maze," Proc. Int. Symp. on the Theory of Switching, Harvard University, Cambridge, Massachusetts, 1–3, (1963).

[63] Peat, Marwick, Mitchell and Company, "New Systems Requirements Analysis Program," UMTA, Technical Development Plan, August 11, 1972.

[64] Peat, Marwick, Mitchell and Company, "New Systems Requirements Analysis Program," UMTA, General Functional Specifications, Work Item 3: Development of Interactive Sketch Planning Techniques, December 29, 1972.

[65] Pratt, R. H., "Demand Forecasting for Short Range and Low Capital Intensive Options," *UTDF Conference*, Williamsburg, Virginia, 1972.

[66] PRC/Systems Sciences Company, "New Systems Requirements Analysis Program," UMTA Executive Summary, UMTA Transportation Planning System Development Project, 1972.

[67] PRC/Systems Sciences Company and A. M. Voorhees and Associates, "A Manual Technique for Preliminary Transit Feasibility Analysis," 1973.

[68] Rapp, Matthias H., "Man-Machine Interactive Transit System Planning," *Socio-Economic Planning Sciences*, Vol. 6, 1972, pp. 95–123.

[69] Roberts, P., "Long Range Travel Demand Forecasting for Urban Transportation Facilities," UTDF Conference, Williamsburg, Virginia, 1972.

[70] Ruiter, E., "Analytical Structures," UTDF Conference, Williamsburg, Virginia, 1972.

[71] Schneider, Norton, "Access and Land Development," HRB Spec., Report 97, 1968.

[72] Schneider, Norton, "Access and Development Prototype Project," High-

way Research Record 293, 1969, pp. 147–154.

[73] Urban Mass Transportation Administration Administration," "New Systems Requirements Analysis Program," UMTA Transportation Planning System (UTPS), Course Notes, October 3, 1972.

[74] Urban Mass Transportation Administration, "New Systems Requirements Analysis Program," UTPS Reference Manual, 1973.

[75] Wachs, M., "Social, Economic and Environmental Impacts,' UTDF Conference, Williamsburg, Virginia, 1972.

[76] Weber, A., "Über den Standort der Industrien," Tubingen, Germany, 1909.

[77] "Multi-Mode Transportation Model, A Summary Description," MITRE, Report No. M71-24, June 1971.

Chapter 8

Models in Law Enforcement and Criminal Justice

By

Saul I. Gass

SUMMARY 233

I. INTRODUCTION 235

- The Criminal Justice System 235
- Law Enforcement Decisions 238
- Court Decisions 242
- Corrections Decisions 242

II. LAW ENFORCEMENT MODELS 244

- Models for Manpower Resource Allocation 244
- Patrol Beat Design 251
- Crime Prediction and Random Patrol 254
- Simulation of the Police Emergency Response System 256
- Perspective of Law Enforcement Models 261

III. COURT MODELS 262

- Court Statistics 262
- Court Planning and Operations 263
- Court Simulation Models 266

IV. CORRECTIONS SYSTEM MODELS 268

V. HOLISTIC APPROACHES TO ANALYSIS OF THE LE/CJ SYSTEM 269

REFERENCES 271

Models in Law Enforcement and Criminal Justice

SUMMARY

Important and persistent decision-making problems inherent in the administrative and operational contexts of this country's law enforcement and criminal justice agencies are now being resolved by the new methodologies of modern scientific decision making. In this chapter, we describe many of these problems and the methodologies which have been applied to solve them. We do not mean to imply here that a specific problem as interpreted by one agency and that agency's approach to solving it can be applied universally. There are, however, certain classes of decision problems which have been studied by the law enforcement and criminal justice research and operational communities which do have general structures and approaches for resolving them. Hence, this chapter discusses such problems within the framework of classes of models or decision areas (see Summary Table).

As we are concerned with all of the components which combine to form the law enforcement and criminal justice (LE/CJ) system, we first describe the general structure of that system and note that it is more an amorphous structure than a true integrated system. We can, however, subdivide the LE/CJ by decision areas and we next describe the range of decisions faced by administrators and operational personnel in law enforcement, courts and corrections. Emphasis throughout is on law enforcement as this area has been subjected to a more intensive review (and has received more funds) than the other areas.

Many basic law enforcement problems are concerned with the determination of the number of personnel (or patrol units) to field during a particular shift as a function of the expected number of calls for service. We review statistical procedures for forecasting calls for service and models which take these forecasts and develop manpower loadings which satisfy specified criteria, such as immediate assignment of a unit to a call for at least 85% of the calls. Other related decision models which are discussed deal with the division of a city into patrol beats of equal workload and the use of statistical procedures for crime prediction and randomizing the patrol pattern of a patrol unit.

Major projects are underway to develop models which simulate the operations of the police dispatching and patrol activities. Such models will enable planners to evaluate proposed changes to these areas prior to any field test, and thus be able to select more viable alternatives for implementation. These models are reviewed, along with a discussion of their future role in law enforcement planning.

Court models and decision problems revolve about the use of statistical data for analyzing administrative and scheduling (calendaring) problems, and the allocation of resource models to increase the number of cases processed in a given time period. Simulation models have been effective here in evaluating proposed structural changes, e.g. increase in the number of grand juries, and procedural changes.

Work in the corrections field has been limited to models as an aid to sentencing and the selection of proper treatment of individuals, and as a means

Preceding page blank

LAW ENFORCEMENT AND CRIMINAL JUSTICE

Models Discussed in Chapter

Summary Table

Model Decision Area	General Type	Important Characteristics
Law Enforcement:		
Resource Allocation (St. Louis **LEMRAS**)	Forecasting and Queuing Model	Uses historical data to forecast level of crimes by type and determines number of patrol units required to maintain a desired level of response
Resource Allocation (NYC RAND)	Dynamic Programming	Predicts the performance of a specified manpower level and police deployment in terms of dispatch delay, average response time, patrol frequency and time available for patrol
Patrol Beat Design	Linear Programming	Divides a geographical area into specified number of patrol beats, each with approximately same level of workload
Crime Prediction	Multiple Regression, Exponential Smoothing	Predicts crime levels by type and area: determines if crime factors are at specified levels and a crime can be expected to occur in an area
Random Patrol (Edina)	Monte Carlo, Randomization	Assigns units to patrol areas based on probability of crime occurrence
Emergency Response (District of Columbia **MATHEMATICA**)	Simulation	Simulates movement of a patrol force for evaluation of alternative decision rules for dispatching calls for service, assignment of patrol units and general patrol strategies
Courts:		
Allocation of Resources	Queuing, Linear Programming	Determines number of judges to serve expected number of cases; allocates cases to judges over time to maximize number of cases cleared
Court Operations	Regression Analysis, Simulation	Evaluation of specific alternatives for speeding up and improving court operations
Court Calendaring	Computer-Assisted Logical Evaluation, Simulation	Develops calendar of cases so as to reduce delays, over or underloading, and waste of personnel time
Juror Selection, Juror Waiting Time	Simulation	To determine correct size of jury pool to maximize utilization
Corrections:		
Sentencing, Rehabilitation	Statistical Analysis	Evaluation of different correctional programs on recidivism
LE/CJ System:		
Offender Flow and Cost (Science and Technology Task Force/**JUSSIM**)	Steady-State, Probabilistic Linear Flow, Simulation	Evaluation of parametric changes and alternatives in the LE/CJ system in terms of total system cost, manpower, criminal population

of analyzing correctional and rehabilitative programs.

Attempts at analyzing the interrelationships of the components parts of the LE/CJ have produced models which can be employed by State and jurisdictional planners to aid in evaluating total LE/CJ programs and their cost. Such holistic models have also been used to demonstrate to LE/CJ planners how their decisions and particular activities impact the rest of the LE/CJ system.

MODELS IN LAW ENFORCEMENT

Trial by Mathematics

In 1964 a black man and his white wife were convicted of a mugging in San Pedro, California mainly because they were the only couple in the area who matched the reports of witnesses on five counts: the girl was a blonde, she had a ponytail, her companion was black, he had a beard, they drove a yellow car. The prosecutor estimated the probability of each characteristic separately —1/10 for the car being yellow, 1/1000 of the companion being black, and so on. He multiplied the five fractions and convinced the jury that the probability was 1/12,000,000 that a matching couple lived in the vicinity. Not until four years later did the California Supreme Court reverse the decision after a judge less ignorant of mathematics persuaded the court that the estimate should have been 41/100.

(News item, *Time*, April 26, 1968, p. 41, as reported in *Scientific American*, October 1972, p. 111).

I. INTRODUCTION

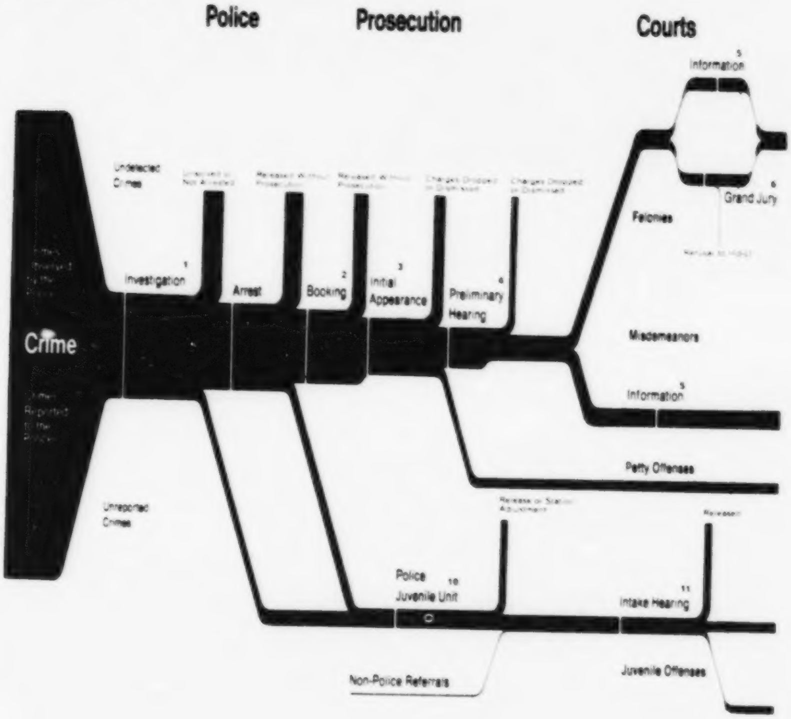
The above news item serves to illustrate how the use of mathematical analyses and related decision-making models in the areas of law enforcement and criminal justice (LE/CJ) differs, in a sense, from the use of decision-making models in most other fields. Although, as we shall see, the majority of LE/CJ models do not deal with a specific individual or criminal situation, the decisions made with the help of such models have a more direct bearing on our daily lives than is the case in other fields. Thus, it is felt that humanistic considerations must always be borne in mind when considering a LE/CJ model—our models are not infallible and rest heavily on assumptions and available data.

Until the 1960's most LE/CJ models were of the simple probabilistic variety dealing with the probability associated with a particular circumstance or set of circumstances. If an impressively small probability holds, then some contend "that 'beyond a mathematical doubt' transcends 'beyond a reasonable doubt' in the court room" [1]. We shall not, however, dwell on this restricted area of criminalistics as more important and exciting advances in LE/CJ models have recently occurred and are becoming an influential part of the LE/CJ decision-making framework. We shall review the background and development of such models in this chapter.

The Criminal Justice System

We tend to use the phrase "Criminal Justice System" as if it meant the dictionary definition of "a regularly interacting or interdependent group of items forming a unified whole." Even a cursory investigation shows that the "unified whole" is an illusion. As noted in [2], "every village, town, county, city and State has its own Criminal Justice System and there is a Federal one as well. All of them operate somewhat alike. No two of them operate precisely alike." The Criminal Justice System has three separately organized parts —police, courts and corrections—each having distinct tasks. Figure 1 shows a general view of the Criminal Justice System which is more or less applicable to most jurisdictions [2]. There are over 46,000 public agencies in the Criminal Justice System [4].

Figure 1 can also be viewed as a parital decision-making framework or tree for LE/CJ, where a branch represents someone's decision that has to be made, e.g. make an arrest, plead innocent, schedule trial, determine length of sentence. We note that there are many decisions which are not included in Figure 1, especially those dealing with the operations and planning of the LE/CJ components. In what follows, we do not mean to catalogue the full range of LE/CJ decisions nor the full scope of how models have been of assistance. However, we shall describe by major components—law enforcement, courts, corrections—many decision situations, to be followed with discussions of how specific models have been applied to aid the LE/CJ decision process. Due to our tradition of assigning most law enforcement and the administration of justice activities to the States, we shall emphasize State and local problems over Federal concerns, recognizing that


This chart seeks to present a simple yet comprehensive view of the movement of cases through the criminal justice system. Procedures in individual jurisdictions may vary from the pattern shown here. The differing weights of line indicate the relative volumes of cases disposed of at various points in the system, but this is only suggestive since no nationwide data of this sort exists.

Police
Prosecution
Courts
Crime
Undetected Crimes
Unsolved or Not Arrested
Released Without Prosecution
Released Without Prosecution
Charges Dropped or Dismissed
Charges Dropped or Dismissed
Investigation 1
Arrest
Booking 2
Initial Appearance 3
Preliminary Hearing 4
Information 5
Grand Jury 6
Refusal to Indict
Felonies
Misdemeanors
Information 5
Petty Offenses
Unreported Crimes
Release or Station Adjustment
Released
Police Juvenile Unit 10
Intake Hearing 11
Non-Police Referrals
Juvenile Offenses

1 May continue until trial.

2 Administrative record of arrest. First step at which temporary release on bail may be available.

3 Before magistrate, commissioner, or justice of peace. Formal notice of charge, advice of rights. Bail set. Summary trials for petty offenses usually conducted here without further processing.

4 Preliminary testing of evidence against defendant. Charge may be reduced. No separate preliminary hearing for misdemeanors in some systems.

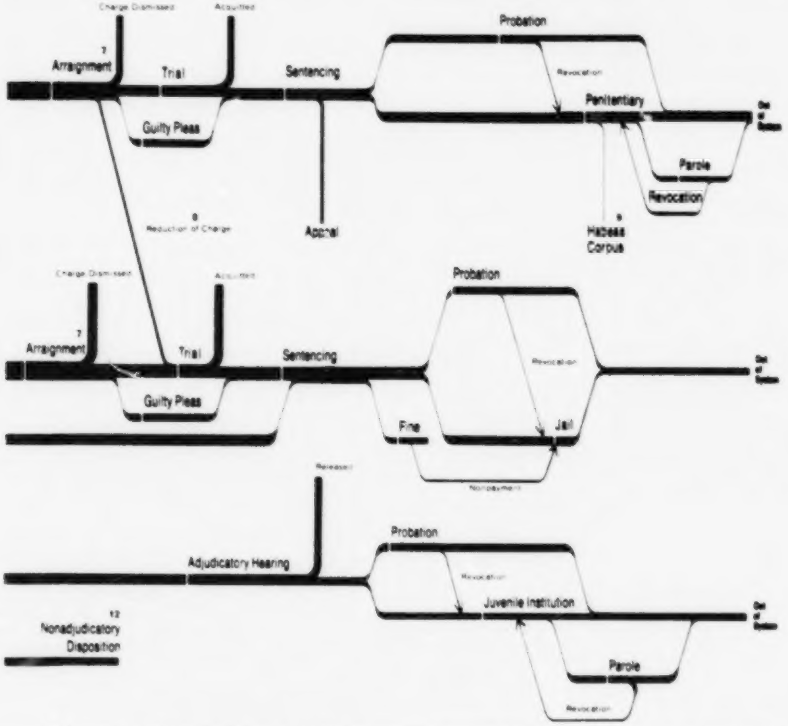
5 Charge filed by prosecutor on basis of information submitted by police or citizens. Alternative to grand jury indictment; often used in felonies, almost always in misdemeanors.

6 Reviews whether Government evidence sufficient to justify trial. Some States have no grand jury system; others seldom use it.

FIGURE 1—*A general view of The Criminal Justice System*

there is some overlap, especially in the areas of courts and corrections.

We first note that even though the LE/CJ system consists of operating entities which are administratively independent and, in some situations, even have conflicting objectives, we can analyze the behavior of the system, especially in terms of how changes in the operation of a major component impacts the ability of other components to perform their functions. For example, about 1.8 million persons were arrested for drunkenness in 1971 [5]. If new approaches to the handling of this problem were instituted on a large scale, e.g. detoxification and treatment centers, then this would change not only the time the police would have for performing other enforcement activities, but the operations of the related courts and jails would also be modified. (The Commonwealth of Massachusetts now considers drunkenness as an illness, not a crime.) Changes like new approaches to correctional institutions, new sentencing and parole procedures and other proposals cut across the components of the LE/CJ system and cannot be evaluated in isolation. A system analysis and related model of the total LE/CJ system could enable planners to measure such impacts and decisions. We shall discuss such models below.


Charge Dismissed
Acquitted
7
Arraignment
Trial
Guilty Pleas
Sentencing
Probation
Revocation
Penitentiary
Out of System
Parole
Revocation
8
Reduction of Charge
Appeal
9
Habeas Corpus
Charge Dismissed
Acquitted
7
Arraignment
Trial
Guilty Pleas
Sentencing
Probation
Revocation
Fine
Jail
Nonpayment
Out of System
Released
Adjudicatory Hearing
Probation
Revocation
Juvenile Institution
Out of System
Parole
Revocation
12
Nonadjudicatory Disposition

7 Appearance for plea; defendant elects trial by judge or jury (if available); counsel for indigent usually appointed here in felonies. Often not at all in other cases.

8 Charge may be reduced at any time prior to trial in return for plea of guilty or for other reasons.

9 Challenge on constitutional grounds to legality of detention. May be sought at any point in process.

10 Police often hold informal hearings, dismiss or adjust many cases without further processing.

11 Probation officer decides desirability of further court action.

12 Welfare agency, social services, counselling, medical care, etc., for cases where adjudicatory handling not needed.

Source: Reference [2]

Law Enforcement Decisions

For our purposes we are concerned mainly with the operation of an urban police department. However, much of our discussion is independent of the size of a department and its demographic descriptors. As we take a close look at what we define to be the set of law enforcement agencies, we find extreme diversity to be a major characteristic. A Department of Justice survey [4] noted that there are over 14,000 separate agencies responsible for enforcing laws on the Federal, State and local levels of government. Most of these are located in boroughs, towns and villages; with about 3,000 each in the counties and cities. Their size ranges from the one man police department of Alpha, New Jersey to the 33,000 man department of New York City.

From an organizational point of view, each police department has a full range of decision-making problems, many of which are similar to those encountered in industrial or other governmental settings. We next describe some of these problems in law enforcement terms.

Much information is available which relates the level of crime with demographic and socio-economic data [6]. Our approach to measuring the effectiveness of our law enforcement agencies ranges from the empirical measure of the latest crime statistics, to the emotional measure supplied by the latest rash of unsolved murders. There does not appear to be an agreed upon measure. Unlike profit-oriented organizations, the activities of a police department in terms of service to the community cannot be measured in dollars or even in terms of a more general measure. (This is, of course, true for most governmental organizations.) For the police, this problem is further compounded by the lack of consensus on what the police should be doing, i.e. what is their job?

The problem of police activity and measures was studied in the report [7]. The authors were concerned with decision-making in the area of police patrol. They state that "the major functions and activities of all police patrol forces are (1) to prevent and deter crime, (2) to apprehend criminals, (3) to respond to calls for assistance from the public, and (4) to regulate certain noncriminal activities such as traffic." They discuss problems in terms of efficiency, effectiveness, and equity. Efficiency deals with measures internal to the system (the number of patrol vehicles operating during a shift), effectiveness implies measuring output or external effects (the level of crime in a city), and equity has to do with how a service and its benefits are distributed among the population. For our purposes, we shall use the general term of measure of effectiveness for any criterion which aids the decision-maker to select among alternative actions.

In addition to the level of crime by crime types, other law enforcement measures of effectiveness include the average response time to a call for service, the number of arrests, the number of crimes which are cleared (an arrest or other disposition for a crime), the number of convictions, the rate and direction of change of the level of crimes, the workload for a patrol unit, the number of calls for service made to the police. We might also consider such subjective indicators as how the public "feels" in terms of safety in streets, as measured by a survey. For example, a Gallup Poll [8] taken in December 1972 resulted in 21% of the urban respondents stating that crime was their community's worst problem, and that fear to walk alone in their neighborhood increased from 31% in 1968 to 49% in 1972. Thus, one of the key open questions for this part of the LE/CJ system is how can we determine if the decisions and actions made by law enforcement agencies improves their effectiveness?

From a model point of view, the way around this dilemma (which is not restricted to the LE/CJ area) is to offer, for a particular model, a specialized measure of effectiveness which enables the decision-maker to select among available alternatives. The measure represents a surrogate measure which is both understood and accepted by the decision-maker and is more tractable to formulate within the model framework, e.g. the average response time for a set of new patrol beats instead of the change in the crime rate which results due to the new beats. This is the approach which is used for law enforcement models and other components of the LE/CJ system. In many instances, a single measure cannot be used and a composite measure of effectiveness, which could be a set of numbers or set of information, must be employed in making the decision.

Like most other organizational units, law enforcement decisions can be dichotomized into administration and operations. We next look at some of these decision situations stated in law enforcement terms.

In the domain of administration we find the police faced with the following problems:

1. For given demographic parameters —population, level of crime, urban configuration, etc.—how many full-time employees should a police department have and how should they be distributed between patrol, detective and administrative divisions? We find, as a rule of thumb, that the police force in large urban areas consists of about 4 employees per 1000 population and ranges down to about 1 employee per 1000 population in the small cities and towns.

2. Given a force level, how should the manpower be allocated to police districts, precincts, shifts and patrol beats over time?

3. How should a city be divided into police districts, precincts or beats based on demographic, geographic and crime parameters?

4. How should a police district be broken down into individual police patrol beats so that each patrol unit has equal work level? What definition of work do we use?

5. How should the patrol force be divided between a one or two-man car, foot or scooter patrols and which patrol type of types assigned to a particular beat?

6. In the long-range planning area we need to determine manpower levels, location of new facilities, contingency riot and other emergency plans, equipment needs, etc. How can we determine the value of technological advances in car locators, nonlethal weapons, patrol vehicle design, communications?

In the operational area we have the following problems:

1. How to identify a particular pattern of crime which is related to the same criminal or set of criminals?

2. Given that we have identified a particular crime pattern, how do we identify a probable set of suspects?

3. Given someone who has been arrested, how can we determine which crimes (if any) are associated with the suspect?

4. With regard to the geographic distribution of crime, how can we best pre-position special patrols in order to increase the probability of apprehension and to increase the patrols' deterrent power? How much preventive patrol should a patrol put in? Is preventive patrol effective in deterring crime?

5. How should the limited investigative effort be allocated and to which crimes? (A preliminary study for the President's Crime Commission [9] reveals that the investigative force usually works on those crimes which have a solid clue, e.g. named suspect, auto license plate number, and do little Sherlock Holmes type work on the run-of-the-mill crime.)

6. Given a call for police services, which unit should be selected to respond? As police forces respond to such calls, how should the remaining forces be dynamically re-allocated based on current and expected demands?

To obtain a better understanding of specific law enforcement problem areas, we next discuss the major set of activities of police agencies, termed the police emergency response system. As background, we note that police activities were first scrutinized in detail in light of modern systems analysis and modeling procedures by the staff of the 1965 President's Crime Commission. The technical studies done by the Crime Commission were reported in [9] by the Science and Technology Task Force. The Task Force modeled the functions of a generic police department in terms of a flow chart of activities associated with an occurrence of a crime. This chart is shown in Figure 2. (The source is [7] which modified the Task Force Chart to include the noncrime activities of the police.) The response process begins with the detection of a crime by a patrol unit or by a report to the police by an alarm, a witness or a victim. After receiving the information, an appropriate response must be made and patrol officers are dispatched to the scene. This is followed by search and investigation (interrogation, data gathering, suspect checking) and then, perhaps an arrest.

To support the emergency response process, decisions must be made and activities performed, some of which are based on analysis and some of which are based on experience and intuition. At this time, most such decisions are made without the aid of formal models. (Here, we include decisions required in the planning function as well.) We do note that crime detection and prevention activities have used pin maps—which in some agencies have been replaced by computer contour and density maps—and basic statistical analyses, e.g. the number of crimes which occurred in the preceding day, to aid them in prepositioning forces and developing patrol routes and patrol beats.

The main measure of effectiveness of the response system is considered by most analysts to be the patrol response time, i.e. the time from receipt of a call for police services until the arrival of a patrol unit at the scene [7], [9], [10]. A Crime Commission study in [9] showed that for emergency calls for service for which an arrest was made, the average response time was 4.1 minutes; for those emergency calls which did not culminate in an arrest, the average response time was 6.3 minutes. A similar pattern existed for nonemergency calls for service. A short response time also has the benefit of improving the public's view of the department and thus, the public's measure of effectiveness. A department, concerned with response time, can make appropriate changes in its mode of operation, each such change representing a decision which has to be evaluated in terms of improved effec-

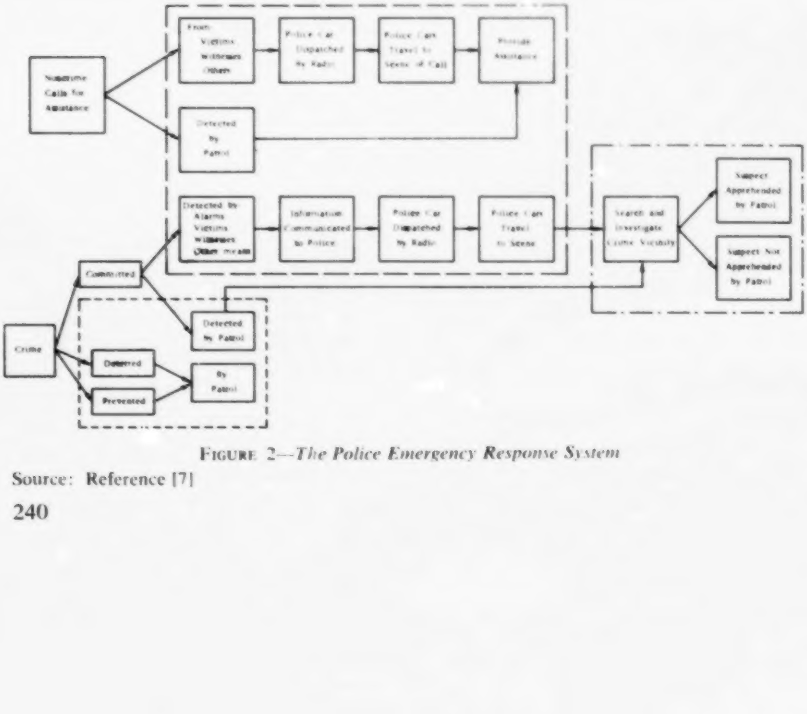

Noncrime
Calls for
Assistance
From:
Victims
Witnesses
Others
Police Car
Dispatched
By Radio
Police Cars
Travel to
Scene of Call
Provide
Assistance
Detected
By
Patrol
Detected by
Alarms
Victims
Witnesses
Other means
Information
Communicated
to Police
Police Car
Dispatched
By Radio
Police Cars
Travel
to Scene
Search and
Investigate
Crime Vicinity
Suspect
Apprehended
by Patrol
Suspect Not
Apprehended
by Patrol
Committed
Crime
Detected
by Patrol
Deterred
Prevented
By
Patrol

FIGURE 2—*The Police Emergency Response System*

Source: Reference [7]

tiveness and associated costs (personnel, equipment). To further illustrate the law enforcement decision problem area, we next describe the elements of the emergency response process in more detail. The reader will be able to note those elements which can be changed so as to improve the total response time. Some of the decisions to effect such changes can be supported by decision models. We shall discuss a few of these models in the sections that follow.

Once a crime has been discovered, the information has to be relayed to the police communications center. Delays can occur if there are not enough trunk lines or operators. The operator must determine the patrol beat location of the call and pass the request to a dispatcher. Some departments combine the roles of operator and dispatcher. For example, Chicago has organized its communications center so that a call made from any police district is automatically assigned to an operator/dispatcher who only handles that district and is quite familiar with the district's streets and patrol beats. New York City has installed a computerized system which automatically matches the location of the scene with the patrol beat and displays this information to the dispatcher. The use of a special police emergency number (911) and the ability to dial that number from a coin box without using a coin has been instituted in some localities. Special alarm boxes for voice communication directly to police and fire dispatchers are being evaluated, while silent alarm and telephonic alarm systems tied directly to the police communications center are in use.

The police dispatcher must select an available patrol unit to respond to an emergency call. In many instances, the beat car is busy and a neighboring free car is assigned, or the call is stacked in a queue depending on its priority. Each department develops unique dispatching rules for the assignment of an emergency call for service to a patrol unit. Decision considerations include the priority of the call (robbery in process, drunk in the street), which unit appears to be closest in terms of travel time to the scene, availability of the corresponding beat car, the preemption of a car servicing a lower priority call, one-man or two-man unit response. Studies have been made to measure the decrease in average travel time based on variations in the dispatching rules. For example, the dispatcher does not know which patrol unit can get to the scene the fastest. Car locator systems (which can be rather costly) have been discussed in terms of supplying this information to the dispatcher.

The ability of the dispatcher to reach the selected patrol unit depends on the congestion level of the radio communications channels. Many departments find the radio waves overloaded in busy periods and much time is wasted before a link is established. The analysis of congestion in telephone and radio systems can be done by models which indicate expected delays and what decrease in the delay we could expect by addition of a line or operator. As digital information can be packed tighter than voice over the same line, digital printers are being experimented with for the transfer of certain information.

Once the unit arrives on the scene, the law enforcement personnel are then faced with many related decision problems. These include the type of search conducted, initiation and carrying out of an arrest, the reassignment of the patrol unit to another call or to preventive patrol, evaluation of evidence, suspect determination, the final clearance of a crime.

From the above discussion the reader should now have some understanding of the complex process associated with the planning and operations of a typical urban police department. Critical decisions ranging from what next year's manpower level should be to how many men should be on duty the next shift are being made on a routine basis. Some of the more important models which have been developed to aid law enforcement personnel in making some of their decisions are described in a following section.

We continue our discussion of the LE/CJ decision process by next reviewing the court system decisions and the correctional system decisions.

Court Decisions

The country's court systems are quite diverse with over 13,000 State and local courts [4]. Not all courts are criminal courts, e.g. probate and family relations courts. About 1,700 courts are located at the State level of government, while the remaining are at a local level (county, city, township). The analysis of court related decisions and the application of models to aid in these decisions is compounded by the lack of central authority. As the President's Crime Commission pointed out: "The complex problems of court administration will not yield to any one simple solution, but a well-structured and efficiently organized system is a condition precedent to further change. Rebuilding the structure and organization of the administration of criminal justice has two aspects, the creation of a unified, simplified court structure within a State and the establishment of clear and direct administrative responsibility within that system [11]."

The decision framework of a court system, as distinct from those made by a judge in a particular trial, concerns the administration and day-to-day management of the court's facilities and personnel. This, as we shall see, has been the main direction of application for models in this area. A basic concern of a court system is the reduction of the delays in the processing of a case. The major delays occur at the following key points in the judicial process: arrest to first judicial appearance, appearance to formal charge, formal charge to pretrial proceedings, pretrial to trial, conviction to sentencing, sentencing to appellate review. A basic question here is to determine what changes in a particular court system will be effective in reducing delay, recognizing that such changes call for financial, personnel and even legal modifications of the present system. Some decisions to have more judges and juries, or to make procedural changes aimed on delay reduction can be evaluated with the aid of models.

Other court management problems include those of calendaring or scheduling the cases for a particular date, the assignment of a judge and the finding of an available courtroom, and proper utilization of jurors. Measures of effectiveness for a court system could include the following: the average elapsed time of court proceedings for each type of case; holding the number of judges and courtrooms to a minimum while the elapsed time for each case is kept within certain limits: reduction of the backlog of total cases; or some measure of the social good of the court process, i.e. the constitutional guarantees granting an accused man the right to a speedy trial [12]. The minimization of the average delay time for all cases might be an appropriate surrogate measure.

Contrasted to the decision problems in the management and administration of a court system is the set of decisions related to the disposition of a particular case. These include pretail release of a defendant, with or without bail; the granting of probation, the type and length of sentence; and the assignment to a particular correctional institution or process. We would hope, based on the facts of a case and the particulars of a guilty party, that the sentencing process would be such to increase the chances that the guilty person would not become a recidivist. How to measure this relationship has proven difficult and little work has been done in this important area. An additional area is the use of appropriate models by prosecuting and defending attorneys.

Corrections Decisions

The corrections section of the LE/CJ system is, like the other components of law enforcement and courts, basically a State and local function. Of the over 7,600 corrections agencies, there are

4,435 adult correction agencies, 724 juvenile correction agencies, and 2,445 probation offices under State and local administration [4]. Over one million offenders are assigned to the above agencies at a cost of over one billion dollars [13]. Like the court system, the correction system decisions can be dichotomized into (1) management and administration decisions and (2) decisions affecting an individual offender. As noted in the President's Commission Task Force Report on Corrections [13], both areas are handicapped by the problems of data availability and relevance: the data essential to the making of sound decisions often are not available, and information that is available may be irrelevant to the outcomes which determine whether the decision was sound. There is a major need to determine what data are necessary, for example, to decide to parole an offender, or to institute a new type of rehabilitation program.

It is generally agreed that the established goal of today's corrections system in the United States is given by "Reformation, not vindictive suffering, should be the purpose of penal treatment" [13]. This is contrasted to the deterrence theory of correctional institutions. The reform approach has led to a wide range of services and specialized institutions directed to reforming the offender: vocational training, prison schools and factories, probation and parole, community programs, halfway houses. "These and similar measures have never been tested definitively, and the reform movement has led an uneasy coexistence with the deterrence and earlier harsher approaches" [13]. How to evaluate such different approaches to corrections is a viable area for the application of certain models. Similarly for the decision areas associated with the course of an offender within the correctional environment. This latter area is an extremely difficult one as it revolves about information which is basically qualitative and open to wide interpretations. The Corrections Task Force noted [13]:

> "Whether the decision is to invoke the judicial process, to choose between probation or imprisonment, to select the appropriate degree of security in a correctional institution, or to determine the timing for release from incarceration or the necessity for revocation of parole, judicial and administrative decision-makers are concerned with very similar issues:
>
> 1. The degree or extent of threat to the public posed by the individual. Significant clues will be provided by the nature of the present offense and the length of any prior record.
> 2. The nature of the response to any earlier correctional programs.
> 3. The kind of personal stability and responsibility evidenced in his employment record, residential patterns, and family support history.
> 4. The kind of personal deficiencies apparent, including educational and vocational training needs.
> 5. The personal psychological characteristics of the offender that determine how he perceives the world and his relationship to it."

What measures of effectiveness can we apply to the correctional process in the evaluation of the success of the process or in the selection of alternative solutions to the decision problems? On an individual basis, the obvious measure is that the offender, upon his normal return to society, does not commit any additional crimes. Variations of this theme include the expected proportion of the current convicted population that will be in prison as the result of current or future convictions, the expected number of times current prisoners will be reincarcerated for new crimes, the equivalent annual cost to society for each criminal career, and based on estimates of first offender input, the expected system population over a 10-year period [14], [15].

For corrections we must emphasize that research and application of models in this major element of the LE/CJ system is extremely weak.

II. LAW ENFORCEMENT MODELS

Models for Manpower Resource Allocation

The most important task of police operations is the proper allocation of resources to meet the expected demand for service. Here we are basically concerned with the level of the patrol force on a daily or even shift basis and a department's desire to respond to emergency calls for service within a specified time.

There is a related planning problem of determining the level of the patrol force and total personnel for the next budget period. Both problems require some means of predicting the future course of events. Such predictions, as we shall note over and over again, must be based on reliable and consistent data gathering procedures.

We shall discuss the resource allocation problem by first discussing simple projection procedures and then discussion a major contribution to police resource allocation, that of the resource allocation technique developed by the St. Louis Metropolitan Police Department.

In law enforcement, the use of simple statistical models have a long history. These include the use of pin and other maps to locate crimes by types in an attempt to glean some behavioral crime pattern and predict the location of the next incident or to highlight the dense areas of crime, and graphs and tables showing frequency counts over time which indicate trends and which attempt to convey some measure of police effectiveness. More recently, with the advent of computers, advanced statistical procedures are being employed. We shall discuss some of these applications which deal with the problem of predicting the level of crime in future time periods in specified areas of a city. These types of models are useful in scheduling the total level of patrol forces required for the next period of time and for the shifting of patrol forces from expected low crime areas to high crime areas.

For statistical analyses to be effective, proper reporting procedures must be instituted to insure the reliability of the results. Standard statistical procedures which can be applied by hand, if the data is not too extensive, or which are available for most computers, can offer a full range of analytical procedures, coupled with the proper statistical tests for indicating that the assumptions of the model are correct. The use of these type of analyess for planning purposes also requires that the police department establish basic guidelines which relate to the level of criminal activity to workload for a patrol officer, detective, and other personnel. Given such planning ratios, and statistical predictions based on past events, the department should then be able to estimate the desired total personnel level, as well as the breakdown between the personnel divisions.

For a given patrol force, we next address the problem of how many patrol units should a department have during a particular tour of duty. A very powerful resource allocation procedure has been developed by the staff of the St. Louis Metropolitan Police Department [19], [20], [21]. This computerized system is available from computer manufacturers and others under the name LEMRAS: Law Enforcement Manpower Resource Allocation System [22]. We shall use the acronym LEMRAS in the discussion below.

LEMRAS is designed to (1) predict calls for service for each of a police department's crime reporting areas (police beats or other units) over the near term by eight-hour shifts, (2) the average amount of police time involved in handling these calls, and (3) the expected number of calls which can be handled without delay, i.e. immediately assigned to a free patrol unit in the police district in which the call originates. Thus, a police district commander can determine how many calls he can expect in his area and how many units he should have during a shift which will minimize the number of calls delayed. This measure is a surrogate measure of effectiveness for the implied measure of minimizing the average time to respond. The use of LEMRAS requires that a district commander has flexibility in establishing patrol beats and the assigning of a different number of patrol units during each shift. This is not the case in many police departments as patrol beats and shift manpower levels tend to be fixed, thus removing the ability of the commander and the department to have a more effective flexible response to changing conditions. Also to be more valuable, a police department using LEMRAS should gather its statistics by areas in the city which are smaller than the patrol beats. St. Louis and Washington, D.C. Police Departments, for example, have established such reporting areas.

LEMRAS consists of two main procedures: a statistical procedure for predicting crime level by reporting area and priority of calls, and a queuing model for determining the ability of a police district to respond to its predicted calls, by priority of calls, without delay.

The validity of these models is based on the following assumptions [22]:

1. Police workload, as measured by calls for police service, has a seasonal fluctuation that, within reasonable limits, is repetitive from year to year.

2. The distribution of police workload over the hours of the week is substantially the same for all weeks of the year, and such changes as do occur are slow in developing.

3. Successive event occurrences act as if they are independent.

4. The service times on successive calls for service act as if they are independent.

5. All patrol units in an area being analyzed (such as a police district) are subject to being assigned to answer any call for service in that area.

6. Any calls for service that cannot be dispatched immediately after being received are held by the dispatcher until a unit is available for assignment. Calls are not stacked in the patrol units.

7. Under normal conditions, a unit assigned to one call will not be pulled off that assignment to handle another assignment with a higher priority. Note that this is a "normal condition" and will be violated when necessary.

8. The assignment of the workload for an event (the time required to service the event) can be totally assigned to the hour in which the event is recorded without substantial impact on the analysis of workloads or events.

The statistical method used to predict the level of crime activity is called exponential smoothing [23]. Exponential smoothing is a method of time series analysis which can be viewed as a special type of weighted moving average. The more usual weighted moving averages have a limited period and give equal weight to each month within the period. In exponential smoothing more recent information is given more weight. The weights decrease geometrically, so that after a sufficient number of periods have elapsed, the weights for older data are so small that the older data contributes essentially nothing to the current average.

For LEMRAS, smoothed averages of the number of calls for service and the associated workload are maintained for geographic areas in a city (entire city, districts, beats, reporting areas). Smooth averages are also maintained on the number of calls for each hour of the week and for each week of the year. Using this information, predictions are made by hour of the week for the geographic areas. As noted in [19], the element being predicted are the emergency calls for service which require police patrol response. To perform this prediction the historical data must be collected by week of the year, hour of the week and by geographic areas. This data collection is a major enterprise in itself and requires a strong commitment by the department in terms of personnel and equipment.

As the number of events, i.e. calls for service, have a seasonal effect, the procedure is usually modified to compensate for such weekly and hourly variations. In addition, the estimated calls can be broken into various categories of calls, e.g. crimes against persons, crimes against property, etc. in that each such call has an associated average time for servicing the call. St. Louis predicts for eight classes of radio run emergency calls for service. The final estimates are obtained for the coming week by hours of the week, based on the smoothed observations of the previous weeks. These estimates are then converted into minutes of work required to service the estimated calls for serivce. An example of such predictions is shown in Figures 3 and 4 respectively. For example, Figure 3 shows that over the next three weeks during the Monday hours of 7:00 to 11:00 A.M., a total of 63 calls is expected, while Figure 4 shows that during these three weeks we expect a total of 2,189 minutes will be used to service these 63 calls.

Given these types of predictions for the coming week—that is for each hour of the week we have an estimate of the number of calls for service and the total time it will take to service these calls for each geographic reporting area of the city—we next want to determine how many patrol units must be deployed to maintain a specified level of service. To approach this problem we cast it into a standard queuing model and assume that the pattern of the arrival of calls for service, i.e. the queue input distribution, is a Poisson process, and that the service distribution, which is derived from the distribution of time required to service the calls, is exponential. These are standard assumptions for many queuing situations and appear to be applicable to this problem [19], [20]. The queuing analysis will enable us to determine the number of patrol units which must be assigned to a police district in order to service a specified percentage of the calls without delay. The specified percentage is a command decision and is controlled by the total force resources, available cars and personnel, the department's goals and objectives, and the district commander's analysis of his specific law enforcement situation. For the St. Louis Police Department, the number of patrol units is set so that at least 85% of all calls will be serviced by a patrol unit without delay. Thus, we see that the St. Louis manpower allocation model and the similar LEMRAS model are decision-aiding tools in that they afford the district commander or department administrators some rationale for making patrol assignments over time which achieves a specified type and level of effectiveness.

The assumptions of the LEMRAS queuing model also states that it is a multi-server system, with the patrol

The Ninth District, St. Louis, Mo.

Prediction period—3 weeks starting 03/06/67

(Number of calls per hour)

Day: Hour	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday	Total
7–11	63	66	56	60	60	67	56	427
11–15	83	80	79	78	86	104	78	589
15–19	113	106	110	105	118	133	84	768
19–23	104	96	94	95	122	133	81	725
23– 3	61	55	59	70	118	107	50	518
3– 7	27	24	29	35	50	43	27	235
Day Totals	451	428	427	442	554	587	375	3264

Figures will not necessarily add to total due to rounding.

FIGURE 3—*Predicted Calls for Service*

Source: Reference [19]

The Ninth District, St. Louis, Mo.
Prediction period—3 weeks starting 03/06/67

(Minutes of work per hour)

Day Hour	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday	Total
7–11	2189	2223	1880	2006	2024	2231	1862	14415
11–15	2791	2694	2626	2584	2877	3348	2526	19446
15–19	3669	3426	3569	3398	3863	4201	2662	24789
19–23	3430	3116	3071	3103	3919	4262	2652	23553
23– 3	1974	1807	1898	2255	3786	3423	1635	16777
3– 7	906	803	965	1143	1652	1402	888	7759
Day Totals	14960	14069	14009	14490	18120	18866	12226	160740

Figures will not necessarily add to total due to rounding.

FIGURE 4—*Predicted Work*

Source: Reference [19]

units representing the set of servers. With these assumptions, standard queuing formulas enable us to calculate the percentage of calls served without delay, given a specified number of servers. LEMRAS does these calculations by varying the level of servers and listing for each level the corresponding number and/or percent of calls which are delayed, by priority (class) of calls.

The main formula used in these calculations is known as Erlang's formula and it expresses the probability that the number of calls for service in the system is greater than the level of servers. Using this formula, two tables are generated, Figures 5 and 6, [19]. Figure 5 relates the level of service attained (the number of calls served without delay) to the number of patrol units. This is shown for the St. Louis eight major classes of radio calls. For example, if six patrol units are on duty, we would expect 94 of the total of 169 class 11 radio calls (crimes against persons) to be handled without delay, while almost all calls can be handled at once if there are thirteen or more patrol units assigned to the district. Thus, we see that the table gives the maximum number of units the district commander must assign in order to obtain a specified level of service.

Figure 6 has more operational value in that it shows service levels in relation to the number of units assigned for four-hour periods of a given day. For example, it summarizes the predicted activity in a district for Saturdays during a three-week period, and breaks the activity into four-hour spans, beginning at 7:00 A.M. Across the top of the page are the potential numbers of patrol units, from 5 to 18, which may be assigned. For each number of units we are given, by four-hour periods, the expected total number of predicted calls, the expected number which can be answered immediately, the expected number which must wait to be serviced.

The Ninth District, St. Louis, Mo.
Prediction Period—3 Weeks Starting 03/06/67

Units Event Class	Total	(Number Serviced Without Delay) 5	6	7	8	9	10	11	12	13	14	15
11	169	58	94	127	147	158	164	167	168	169	169	169
21	667	289	432	539	602	635	652	660	664	666	666	666
22	87	36	55	69	78	82	85	86	86	86	86	87
31	10	4	7	8	9	9	9	10	10	10	10	10
41	8	3	5	6	7	8	8	8	8	8	8	8
42	963	328	543	727	839	901	933	949	957	960	962	962
61	82	40	57	68	75	79	81	81	82	82	82	82
71	1279	550	823	1031	1152	1217	1251	1267	1274	1277	1279	1279
Total	3264	1308	2016	2576	2908	3089	3183	3228	3249	3258	3262	3263

FIGURE 5—*Predicted Number of Events Serviced Without Delay Vs. Number of Units Assigned*
Source: Reference [19]

These prediction are for areas in the Ninth District

Prediction period—3 weeks starting 03/06/67

This is for all Saturdays in the prediction period

Units	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Hours 7–11														
Total pred	59	59	59	59	59	59	59	59	59	59	59	59	59	59
No delay	46	53	57	58	59	59	59	59	59	59	59	59	59	59
Delay	13	6	2	1										
Pct no delay	77	90	96	99	100	100	100	100	100	100	100	100	100	100
Pct delay	23	10	4	1										
One more unit	8	3	1	1										
Hours 11–15														
Total pred	92	92	92	92	92	92	92	92	92	92	92	92	92	92
No delay	35	62	77	85	89	90	91	91	91	92	92	92	92	92
Delay	56	30	15	7	3	1								
Pct no delay	38	67	84	93	97	99	100	100	100	100	100	100	100	100
Pct delay	62	33	16	7	3	1								
One more unit	26	15	8	4	2	1								
Hours 15–19														
Total pred	116	116	116	116	116	116	116	116	116	116	116	116	116	116
No delay	3	42	74	94	105	111	114	115	116	116	116	116	116	116
Delay	114	75	42	22	11	5	2	1						
Pct no delay	2	36	64	81	91	96	98	99	100	100	100	100	100	100
Pct delay	98	64	36	19	9	4	2	1						
One more unit	39	33	20	11	6	3	1	1						
Hours 19–23														
Total pred	116	116	116	116	116	116	116	116	116	116	116	116	116	116
No delay	2	39	72	93	105	111	114	115	116	116	116	116	116	116
Delay	114	77	44	23	12	5	2	1						
Pct no delay	2	33	62	80	90	95	98	99	100	100	100	100	100	100
Pct delay	98	67	38	20	10	5	2	1						
One more unit	37	33	20	12	6	3	1	1						
Hours 23–3														
Total pred	93	93	93	93	93	93	93	93	93	93	93	93	93	93
No delay	32	59	76	85	89	92	93	93	93	93	93	93	93	93
Delay	61	34	18	9	4	2	1							
Pct no delay	34	63	81	91	96	98	99	100	100	100	100	100	100	100
Pct delay	66	37	19	9	4	2	1							
One more unit	27	17	9	5	2	1								
Hours 3–7														
Total pred	37	37	37	37	37	37	37	37	37	37	37	37	37	37
No delay	35	37	37	37	37	37	37	37	37	37	37	37	37	37
Delay	2	1												
Pct no delay	94	98	99	100	100	100	100	100	100	100	100	100	100	100
Pct delay	6	2	1											
One more unit	1	1												

Figure 6—*Summary Queue Table*

Source: Reference [19]

the corresponding percentages, and the expected number of waiting calls which would be answered at once if one more unit were addded, e.g., between 7:00 A.M. and 11:00 A.M., it is expected that 6 units could handle 53 of 59 calls (90%), while 6 (10%) would be delayed; by adding one more unit, 3 more calls could be answered at once. It is expected, however, that 9 units could handle all calls with virtually no delay. The table enables the district commander to determine explicitly what the addition or subtraction of a patrol unit will "buy" in terms of changes in the percentage of calls serviced without delay. This is an important and new type of marginal analysis not previously available to police administrators.

The above discussion illustrates how the LEMRAS like models can be used within the planning and operational framework of a police department. The type of decision-making information supplied by LEMRAS is new for just about all police departments. It adds a whole new dimension to police decision making and hopefully, they will be able to integrate this information into their total operations in a manner which improves their effectiveness. This type of manpower allocation model requires a police department to be committed to effecting necessary changes. The LEMRAS analysis can, for example, indicate that to maintain the specified level of service, two patrol units must be added to the present ten patrol beat district. The district commander must determine if he wants to redesign his district into twelve beats or to have the additional two cars act as "floaters." Most departments do not have the flexibility to add or subtract beats as required by LEMRAS—their beat structure is standardized, even to the painting of expensive beat maps in the communications center. Again we emphasize the data needs for these type of models in that most police departments do not gather their data in a manner which can be used directly by LEMRAS. Also, data not usually tabulated are required, e.g. time to service classes of calls for service.

Experience with LEMRAS has shown that it cannot be implemented without concern as to the requirement and needs of the user law enforcement department. For example, it is noted that the use of LEMRAS in Los Angeles caused conflicts with the department's Basic Car Plan, in which a patrol car and personnel are assigned to a particular geographic unit and are expected to spend 50% of the time working with the people of the area.

As final caveats to the use or misuse of LEMRAS, the following comments are noted [22]:

1. LEMRAS does not *predict* exact times and locations of calls for police service. Using standardized statistical techniques, LEMRAS does *forecast* the average number of calls for police service that may be expected to occur in a given area during a given period of time, and the average amount of police time it will take to service that number of calls.

   It should be noted that these forecasts are based upon the normal stability that many observers have noted in police workload data, and that the unusual situations cannot be predicted. If and when these situations occur, it will be necessary for the police administrators to allocate resources on an emergency basis. This caution is included here because it is recognized that such situations may occur.
2. LEMRAS does not relieve field commanders of the need for exercising exceptional judgment in the day-to-day depolyment of manpower, and in evaluating the LEMRAS predictions. LEMRAS is designed and intended to make available to the commander more detailed and complete data on which to base such judgment than has ever been presented to him in the past.
3. LEMRAS does not predict crime nor does it provide guidance for preventive patrol functions. Measurement techniques are limited to called-for services and the manpower required to handle them. The dispatch data collected for LEMRAS is subject to analysis, however, which can provide insight into the development of crime patterns and other conditions which require or justify preventive effort. The form of the preventive effort must be decided by the commander concerned.
4. LEMRAS is not directly related to the measurement of need for manpower in investigative, administrative, or support positions.

In addition to the manpower allocation models described above, there are other models which aid in the evaluation of manpower allocations. These include simulation models (to be discussed below), the deployment analysis and allocation method developed for the Boston Police Department [33] and the New York City Police Department [7], [10], and the work done by the Chicago Police Department [34]. We next present an extract of the deployment model description as given in [7].

That effort involves an analytic model which predicts the performance of a specified manpower level and police deployment in terms of dispatch delay, average response time, patrol frequency, and amount of time available for preventive patrol. The measures used are flexible and any set of criteria which can be analytically related to patrol manpower may be utilized. This method applies dynamic programming as its mathematical allocation technique. It allows the police to determine the manpower necessary to provide acceptable levels of service in terms of each of the measures, and also to allocate a specified patrol force size to neighborhoods, by time of day and week, so as to yield the highest service level obtainable with the given force. This model is more comprehensive than the LEMRAS models in terms of the multiple criteria which may be used, and because it addresses the preventive as well as the response aspects of patrol.

The minimum levels of services provided may be set by the police and community and may vary with the neighborhood or time of day, depending on conditions. The various service level criteria used in this method are flexible but currently include the following: the average number of preventive patrol hours for each outside crime shall not be less than a specified value in each neighborhood, the average frequency of patrol (which is related to outside crime-detection probability) in a neighborhood shall not be less than a specified value, the average time required to travel to the scene of a reported incident or call for service shall not exceed a specified value, and the number of patrol cars for each neighborhood shall not be less than an administratively set minimum.

The use of these measures of service is based on the assumption that additional preventive patrol and reduced response time have a positive effect on crime deterrence and on increasing the chance of apprehending a criminal near the scene of a crime. The existence of such an effect and its actual degree are not known and need to be investigated experimentally. The four measures of service used in the current model are clearly arbitrary and open to revision to suit the local situation.

Procedurally, the number of cars necessary to provide minimum levels of service on each criterion for each district and time period are calculated. The remainder of the force is then deployed (using dynamic programming) to minimize the dispatching delay due to unavailability of free cars.

For this technique the data needed for each command and shift to which police are to be assigned are: area, patrollable street miles, average travel speed of a patrolling car by time of day and day of week, average travel speed of a responding car by time of day and day of week, volume and distribution over time of various types of calls for service, volume and distribution over time of reported outside crimes, and average total service time for each type of call.

The outputs of the deployment analysis and allocation method are: an analysis of the present deployment giving (a) expected response time, (b) expected frequency of preventive patrol, and (c) expected total number of preventive patrol hours; a reallocation of current manpower to satisfy all constraints on levels of service and to minimize delay in dispatching cars due to temporary unavailability of a free patrol car; an analysis of the reallocation of manpower in terms of (a), (b), and (c) above; a determination of the minimum force level necessary in each patrol district to just satisfy each service level constraint.

In a computational exercise using this model and data from New York City, the predicted variation in service levels among precincts was greatly reduced, while average time delay in dispatching due to car unavailability was cut from about 4 minutes to less than 1 minute by reallocating the same city-wide level of manpower. In addition, all measures of acceptable service level were met, whereas they were not all satisfactory when the existing allocation of manpower was used.

Deployment analysis and allocation models have aided decision-making at the highest levels. As noted in [10], such analyses helped New York City to demonstrate the extent to which effectiveness of the patrol force was lower at times of high demand than at times of low demand. This demonstration was a major impetus behind efforts of the mayor and the police commissioner to modify New York State's three-platoon statute and to obtain a fourth platoon, which now operates from 6:00 P.M. to 2:00 A.M. A revision of the law resulted.

The deployment analysis and allocation model described above does represent a conceptual advance over LEMRAS like models due to its multi-objective approach. No operational comparisons between these approaches are available. The manpower allocation models of St. Louis and LEMRAS models are being used in a number of cities. Those law enforcement agencies planning to implement such models would do well to investigate the current and near future state-of-the-art before selecting a particular approach.

Patrol Beat Design

A problem which is related closely to the manpower allocation problem is that of determining the location and size of individual patrol beats. For example, a LEMRAS analysis might indicate that two additional units should be added to a ten patrol beat district. Assuming two more units are available and the police department does have the ability to rearrange beats from time period to time period, how should the district commander divide the district into twelve beats? What measure of effectiveness should be used to determine if the beat designs are acceptable? A number of investigators [24], [25], [26], [27] have written on this subject and there appears to be applicable models to aid in this decision area. To date, no police department has used these models in any concerted manner; the St. Louis Police Department has instituted the most detailed study of this application [25], [26]. The reasons why such models have not been employed include the need for a set of crime statistics by small geographic reporting areas in a city; and the lack of an ability to run a more-or-less sophisticated model on a computer, given the available computer program. We next describe a particular approach to this problem [24]. The reader will note the full set of data required and the interpretive aspects of the model solution.

The basic problem is as follows: given k patrol units to be assigned during a patrol shift, how should the k corresponding patrol beats be determined so that each patrol unit will, on the average, have the same workload? The area of a patrol beat must be contiguous to itself and of such a shape to allow for efficient patrol and response tactics. (A long, narrow beat or star-shaped beat would not be too efficient. Studies have demonstrated the advantage of square-shaped beats [10]. Of course, geographical and political constraints often help shape a beat.)

The International Association of Chiefs of Police (IACP) recommends taking each zone of the city—usually census tracts—and measuring in some sense, the total crime workload in each zone. Then, using heuristics, combining the tracts to form k contiguous beats having approximately the same relative workload. To determine the workload for a tract, weights are given to the various incidents (investigate a criminal act, arrest of a suspect) to yield the weighted workload for a tract. The weights tend to reflect the importance of the incident and the time required by a patrol unit to process the incident. The IACP uses the following weights:

Type of Incident	Relative Weight
Criminal Homicide	4
Forcible Rape	4
Robbery	4
Aggravated Assault	4
Burglary	4
Larceny	4
Auto Theft	4
All Other Offenses	3
Arrests for all offenses except Drunkenness	2
Traffic Accidents	2
Arrests for Drunkenness	1
Miscellaneous Police Services	1

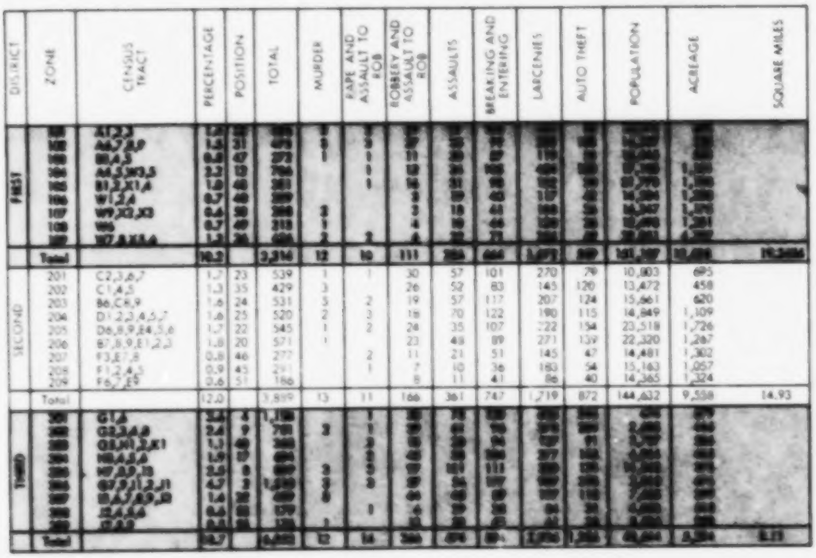
The first seven types of incidents are called Index Crimes as they represent the main categories of crimes reported in the FBI's Uniform Crime Report. Other types of workload measures or "hazard" scores are in use and described in [7] and [10].

One approach to the determination of patrol beats is to attempt to employ analytical procedures for the structuring of the k beats. From the field of operations research we find a situation analogous to the determination of patrol beats, the warehouse location problem. Here, we wish to locate a specified number of warehouses and assign customers to each warehouse such that the total cost of servicing the customers from their assigned warehouses is minimized. The cost could include transportation, delivery and customer relations. This problem definition has recently been extended to include the problem of reapportioning a state into legislative political districts [28]. The problem is to locate a specified number of district centers and assign population units to each district such that the assigned district population must be nearly equal—the one man, one vote concept. The cost, or more correctly, the measure of effectiveness of a set of centers and assignments was taken to be the population moment of inertia, i.e. minimize the sum of the squared distance of each person to his district center. The computational procedure must, along with the measure of effectiveness, allow for the need for contiguous districts, no gerrymanders, and for compact districts (more square than rectangular). We next extend this formulation and associated computational procedure to the structuring of police beats.

Mathematically, the problem can be formulated as an integer programming problem, although the formulation does not necessarily take care of the contiguity requirement [28]. This requirement causes the combination of an analytical procedure (basically solving the transportation model of linear programming) and an automated heuristic adjustment.

To illustrate the process of beat design, we solve the problem by the analytic-heuristic procedure of [28]. The reason for this is that the integer programming model can be quite difficult to solve with available computational tools, although an integer model for beat design is given and solved in [25]. The sample problem to be illustrated next uses data from the 1966 annual report of the Cleveland Police Department as the Department's crime statistics are gathered by census tracts and the beats are formed by combining census tracts. In general, it is recommended that smaller reporting units be used to gather the statistics (as in LEMRAS) and the beats be formed by combining these smaller units. This allows for beats to be designed so that the beat workloads are more equitable. A number of different measures of effectiveness for workload are employed that relate total crime, population, and area of each census tract interpreted as an appropriate moment of inertia.

Again, the basic problem is to combine the census tracts of the City of Cleveland to form contiguous and compact beats. There was a total of $k = 58$ beats and $n = 205$ census tracts to assign to six police districts. Figure 7 shows the number and percentage of index crimes in each of the 1966 beats in the first three districts and the ranking position in terms of index crimes. The reader will note the wide variations in terms of crime rate, population and area of the beats. The computations used five measures of the workload for a census tract: number of index crimes, population, area, level of crime multiplied by the population, and the level of crime multiplied by the area. It is felt that the workload in a patrol beat is some function of these three elements—plus related demographic and geographic data. The computations kept the same six police districts as stipulated by the Cleveland police and attempted to readjust the beats in these districts in a more equitable fashion. There was no attempt to



DISTRICT	ZONE	CENSUS TRACT	PERCENTAGE	POSITION	TOTAL	MURDER	RAPE AND ASSAULT TO ROB	ROBBERY AND ASSAULT TO ROB	ASSAULTS	BREAKING AND ENTERING	LARCENIES	AUTO THEFT	POPULATION	ACREAGE	SQUARE MILES
FIRST	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	
	Total		[illegible]		[illegible]	12	10	111	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
SECOND	201	C2,3,6,7	1.7	23	539	1	1	30	57	101	270	79	10,803	695	
	202	C1,4,5	1.3	35	429	3		26	52	83	145	120	13,472	458	
	203	B6,C8,9	1.6	24	531	5	2	19	57	117	207	124	15,661	620	
	204	D1,2,3,4,5,7	1.6	25	520	2	3	18	70	122	190	115	14,849	1,109	
	205	D6,8,9,E4,5,6	1.7	22	545	1	2	24	35	107	222	154	23,518	1,726	
	206	B7,8,9,E1,2,3	1.8	20	571	1		23	48	89	271	139	22,320	1,267	
	207	F3,E7,8	0.8	46	277		2	11	21	51	145	47	14,481	1,302	
	208	F1,2,4,5	0.9	45	291		1	7	10	36	183	54	15,163	1,057	
	209	F6,7,E9	0.6	51	186			8	11	41	86	40	14,365	1,324	
	Total		12.0		3,889	13	11	166	361	747	1,719	872	144,632	9,558	14.93
THIRD	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	
	Total		[illegible]		[illegible]	12	14	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

FIGURE 7—*Crime by Police District and Beat (Zone)*

Source: Reference [24]

construct beats based on any geographical, political or other considerations.

The partial results of a complete set of computations for District 2 with nine police beats are shown in Table 1. The computational procedure for this district yielded contiguous beats, although there were some minor noncontiguous aspects in a few beats in other districts. For comparison purposes, we show in Table 1 the total number of index crimes in each beat of District 2 for the actual 1966 beats and the beats developed based on equal crime. (Due to different population and unavailability of data, such comparisons cannot be made for other measures.) Table 2 shows for the five measures the amount of the total measure for each of the nine beats.

The decision process in the selection of a patrol beat design is complicated by two factors: there is no agreement as to what measure we should use to construct equal workloads and, since the computational procedures do not necessarily produce a single beat plan with equal workloads, there is no apparent guidance as to which beat plan to implement. To resolve the latter point, we can bring in other measures, e.g. which beat plan causes the patrol units to be out of their beats more often [25]. But this type of evaluation calls for the use of simulation (see discussion below) or advanced mathematical procedures. Even though these factors cannot be resolved to everyone's satisfaction, the ability to change and produce new beat plans readily, based on reason-

Table 1

No. of Index Crimes Per Beat (District 2)

Beats	1966 Actual	Equal Crime
1	539	453
2	429	429
3	531	460
4	520	373
5	545	400
6	571	478
7	277	477
8	291	435
9	186	421
	3,889*	3,926*

*Difference in totals attributed to errors in data preparation.

Table 2

Comparison of Measures

Beats	Crime (C)	Population (P)	Area (A)	$CP/10^3$	CA x 10
1	453 crimes	19306 people	1.159 sq. miles	2010	1145
2	429	16417	1.210	1872	1182
3	460	14530	1.189	2216	942
4	373	19272	1.136	1962	1276
5	400	17864	1.047	1507	1209
6	478	14922	1.017	1960	1123
7	477	20083	1,207	2227	1135
8	435	18604	1.308	1961	1008
9	421	15715	1.234	1800	1256
Desired Average	436	17413	1.167	1946	1142

Source: Reference [24]

able assumptions, is certainly a major step forward for most police departments. The model does not produce a single plan for a given set of data and time period which *must* be implemented. It produces rational plans which enables the cognizant decision maker to evaluate better a range of alternatives based on a wide range of requirements, some of which may be external to the model.

Crime Prediction and Random Patrol

From the above discussions, our ability to predict future levels of crime and police workload can greatly aid a police department in its planning and operational activities. As noted in [29], ". . . strategic decisions relating to force structures and training emphases would be easier if reliable long-range forecasts of crime were available. Tactical decisions on the day-to-day deployment of forces would be eased by accurate short-range forecasts or by forecasts of specific crimes. Analysis of the effectiveness of new (and old) equipment can be performed more accurately if appropriate crime prediction capability exists. An understanding of the relationships between social, economic, political, and moral conditions and crime, expressed in a model, could be used to devise effective crime prevention campaigns. These potential uses have generated interest in the broad technical problem of predicting crime."

The report [29] evaluates a number of prediction models, basically variations of multiple regression and exponential smoothing, as applied to historical data collected for Los Angeles. Quarterly predictions for 27 crime types and 24 geographical areas were computed and these predictions did match the actual crime levels within statistical expectations. Given an ability to predict gross crime levels by geographic area over time (i.e. extrapolative models), we need ways of bringing these predictions into evaluative models for police decision-making. LEMRAS is a good example of this ability. Others are needed, especially causal models, i.e. models which relate actions (alternative solutions) to results and use crime level predictions as inputs. With respect to prediction models [29] concludes: crime statistics can be predicted with sufficient accuracy for some of the possible applications by extrapolation of historical data; the potential applications of crime prediction methodology are sufficiently diverse that no single technical approach is appropriate to all; current qualitative and quantitative understanding of the causes of crime is grossly insufficient to permit the construction of usable causal crime prediction models; extrapolative crime models rely upon the assumption that trends will continue as in the immediate past. For example, changes in public policies (especially, but not exclusively, by police agencies), in economic or social conditions, or in public moral or philosophical attitudes can invalidate this assumption.

Another aspect of crime prediction is the prediction that a crime will occur at a particular time and place. Presently, law enforcement agencies use pin maps, intelligence and other information to deploy stake-outs and patrol specific areas. Is it possible to use prediction models to aid in on-the-scene crime detection and apprehension? For this problem a multiple regression model was tested for the seven index crimes using data collected from the Philadelphia Police Department [30]. The study tested out the relation between crime type and 35 possible related crime factors. Included crime factors were day of week, phase of the moon, temperature, age distribution of population, number and type of schools, etc. The crime factors were used as independent variables of the regression analysis, and the level of crime type as the dependent variable. If the variables were properly related, in a predictive sense, then, for an area of a city, as certain conditions obtained, we would expect a crime to occur. Although the resulting regression equations were not good predictors of crime, the analysis led to further work using cluster analysis, a statistical procedure for grouping of events based on their attributes, here crime factors. This procedure has led to a computer driven query system from which a district commander can receive a report on the likelihood of a crime by type in the district or sectors which he designated. Based on his interpretation of the likelihood value, the commander can then dispatch patrol forces to the area. Evaluation of this procedure is not available at this time.

In general, our ability to predict an event by time and location is not too great. Certainly, given information on the relative frequency of crime and causes of crime, the number of successful predictions should increase. What we need to do is to couple our historical data with a patrol strategy which increases the probability of intercepting crimes in progress. We are, of course, restricted to crimes visible from the street. The concept of random patrol has been introduced to aid in this area. Random patrol is based on the assumptions [31] that

1. irregularity of patrol schedules will for a given patrol rate increase the awareness of the general population that patrol is taking place;
2. randomness of the patrol schedule will discourage the potential criminal who finds that he cannot predict the arrival patterns of patrol vehicles.

Patrolling in a random manner means that there is no fixed sequence by which the patrol visits each point in the area, yet all points in the area are visited within some average time [32].

It is shown in [32] and [10] that the probability of intercepting a crime is rather low, but even so, we should be concerned with how to evaluate alternative approaches to patrol which can increase that probability, without causing deterioration of other measures of effectiveness. A rather detailed mathematical model in [10] addresses the problem "Given the recent pattern of crimes and given limited amount of preventive patrol, how should the patrol effort be allocated along streets to best achieve some object?" The associated analysis finds that the allocation that maximizes apprehension probability depends in a complicated fashion upon recent crimes rates; patrol effort should grow as the logarithm of the crime density and that certain areas, where the likelihood of crime is particularly low, should not be patrolled at all.

An early application of random patrol concepts was based on straightforward Monte Carlo procedures for randomizing the areas visited by patrol units while on preventive patrol. The town of Edina, Minnesota has a 47 man police department and in 1971 had a total of 1,075 index crimes. In 1962, the Edina police, working with personnel from the University of Indiana, initiated the first systematic random patrol activity. The town was divided into four zones, with the four zones divided into a total of 51 subzones. Data on the level of crime were gathered for each subzone and four roulette wheels were designed for each zone. The wheels were divided such that the area of a subzone corresponded to the proportion of crime in that zone. The experiment was run just for the 11:00 P.M. to 7:00 A.M. shift. When a car was free to patrol, the corresponding zone roulette wheel was spun and the car went to that area. It was a very basic system. Folklore about the Edina experiment relates that the first time the wheel was spun it caused the dispatch of the patrol unit to a midnight visit to the city dump. Of course, the officer found someone stripping a stolen automobile. After the first year of operation, burglaries in Edina were down 38 percent on the 11:00 P.M. to 7:00 A.M. shift. Since then much more sophistication has been brought into the system. Under a Law Enforcement Assistance Administration Grant in 1968, the method was computerized and applied on a three shift basis. The guidelines of the computerized Edina random patrol model experiment were as follows [35]:

1. The entire uniformed police force of the Village of Edina, on all shifts, was involved; all were given zone maps.
2. Two of the four zones were patrolled using the random patrol technique; the other two zones were the control zones and were patrolled in the conventional manner. At the end of each week the random patrol and the control zones were interchanged.
3. At the start of each shift, each officer was given a series of random numbers and patrolled the subzones corresponding to the random numbers in the sequence for a period of 15 minutes. After 15 minutes, he proceeded to the subzone corresponding to the next random number on his list. If he received a call, he responded. When the call was completed, he proceeded to the subzone corresponding to the next random number on his list.
4. The field tests proceeded as outlined above for a period of three months.
5. Patrol assignments were made so that, if an officer was in a control zone one day, he would be in an experimental zone the next day.

The Edina project had as its objective the development of a random patrol procedure which would reduce the response of the patrol units to emergency calls for service, i.e. the randomization of the patrol units location based on area crime levels was expected to preposition cars closer to the points of need. A comparison between the year of the random patrol field test with the preceding year showed a 40 percent decrease in average response time from 7.05 minutes to 4.22 minutes; this was achieved even though there was a 19 percent increase in calls for service. It was also noted that while crime in Edina rose 11 percent during the experimental year, crime in the rest of the county rose 17 percent. This difference may be attributed to the random patrol. Another benefit of this type of project is that better data collection procedures resulted.

The Edina work is certainly transferable to other departments, especially those of comparable size. The impact of size and the many major and serious events which can occur in large cities would probably destroy any randomized process for patrolling undertaken by the usual beat cars. However, special preventive patrol units, which do not answer emergency radio calls for service, could probably be used in a fashion similar to Edina to insure a proper level of preventive patrol time in critical areas.

Simulation of the Police Emergency Response System

Up to this point, our discussion of law enforcement decision-making models has been directed towards specific problem areas, with such problems having only a limited, but albeit, important scope with respect to total law enforcement concerns. Granted, at this time we cannot encompass within a single holistic model the complex law enforcement framework; we can, however, attack major components in terms of evaluating how changes in their subcomponents impact one another. And having such an evaluative ability affords us an important aid for selecting from alternative decisions. The main means of accomplishing this for law enforcement, as well as the other areas of the LE/CJ system, has been by the use of simulation models. (However, the work reported by [7], [10], [33] and others, do represent many of the complexities by powerful analytical models. Reference [10] gives a rather complete discussion of a wide range of models to law enforcement problems.)

The functions and the operations of the police emergency response system, i.e. the dispatching and patrol of units to respond to calls for emergency service, was described above and is illustrated in Figure 1. Many aspects of this system can be modeled and such models used to make measured changes, e.g. the determination of the number of telephone lines and operators, the design of beats, the number of patrol units during a shift. These models supply the decision maker with quantitative data by which an alternative policy or operational change can be selected among a set of alternatives. For example, a beat design model will determine a specific structure or a set of equally good structures for a district based on the assumptions and data requirements of the model. There remains an important consideration, however. As every model is limited in its ability to reflect the real world, how can we test out the selected set of beats (solution) prior to actual field operation in order to insure against an unsound solution? An important approach to answering this question, especially for complex systems, is the development of a simulation model which can encompass in a more complete fashion the total system and the interrelationships of the system components, especially the random event nature of a system like the emergency response system.

As the simulation of the emergency response system is rather rich in detail, its execution requires the use of a computer. Such computer simulations have been or are being developed for a number of cities—Boston [36], New York City [7], San Jose [37], Washington, D. C. [38], among others. A detailed description—both technical and nontechnical—of the response system simulation is given in [10]. We shall describe the use of simulation in this decision-making environment as described in [10], [36] and [38].

In a general sense, these simulation models are of value to police administrators in the following ways: they facilitate detailed investigations of operations throughout the city or in part of the city; they provide a consistent framework for estimating the value of new technologies and new approaches to the patrol function; they serve as a training tool to increase awareness of the system interactions and consequences resulting from everyday policy decisions; they suggest new criteria for monitoring and evaluating actual operating systems.

More specifically, the simulation models can be used to test and evaluate concepts such as the following:

1. What changes will result from dispatching rules based on the priority assigned to a call for service? For example, rules which prohibit the dispatch of patrol cars to low priority incidents at times when the system has a high load level can be investigated. Another example might be to divide the patrol force into two groups, where one group deals exclusively with low priority calls on an appointment basis.

2. What will be the effects of modified or completely new patrol beat structures? In other words, can system performance be improved by adjusting beat boundaries, by eliminating beats altogether and assigning territorial responsibility to small groups of cars, by overlapping beats so that any given area continues to receive preventive

patrol even when many patrol cars are servicing calls?

3. What will be the impact of adding or subtracting patrol units? Using more sophisticated performance measures, e.g. requiring that a given percentage of the highest priority incidents receive a response within a set time limit, the number of units required to patrol an area may not equal the number as indicated by present allocation schemes.

4. Which analytic procedures for allocating manpower are in fact "best" for a city? Recent analytic models, e.g. LEMRAS, have advanced a number of promising techniques that should allow improved tour scheduling and deployment of patrol manpower. The simulation model will permit evaluation of these schemes for possible adoption.

5. What impact will new technologies have on the patrol functions, e.g. car locator systems? Such investigations will answer questions regarding the locator resolution desired, the impact on dispatch decisions, and the improvements in total operational performance which may be expected.

6. Are there viable techniques for dynamically repositioning patrol units before and after servicing a call? At times, a rash of calls will draw patrol units to a small area, leaving large gaps in geographic coverage. Is it feasible to reposition the remaining units so that future service is maintained at a desired level, given the present conditions?

In addition to these basic questions, further possible investigations include saturation patrol, the interactions between radio-cars, scooters and foot patrol, the effect of computerizing some aspects of the dispatch function, and the application of random preventive patrol strategies [38].

The following description of the simulation model approach is taken from [36], augmented by material from [38]. The model is designed to study two general classes of administrative policies: the patrol deployment strategy and the dispatch and reassignment policy. The patrol deployment strategy determines the total number of patrol units, whether units are assigned to non-overlapping sectors, which sectors constitute a geographical command, and which areas are more heavily patrolled than others. The dispatch and reassignment policy specifies the set of decision rules the dispatcher follows when attempting to assign a patrol unit to a reported incident. Included in the dispatch policy are the priority structure, rules about inter-district dispatching, the queue discipline.

The simulation models use several important measures of operational effectiveness. These include statistics on dispatcher queue length, patrol travel times, amount of preventive patrol, workloads of individual patrol units, the amount of inter-district dispatches.

The simulation models under consideration work in the following way. They are organized to reflect the spatial relationships inherent in patrol operations, as well as the sequential time nature of events. Incidents are generated throughout the city, distributed randomly in time and space according to observed statistical patterns. Each incident has an associated priority number. As each incident becomes known, an attempt is made to dispatch a patrol unit to the scene of the incident. In attempting this assignment, the simulation logic is structured to duplicate as closely as possible the decision-making logic of an actual police dispatcher. In certain cases this assignment cannot be performed because the congestion level of the force is too high. The incident report then joins a queue of waiting reports. The queue is depleted as patrol units become available [36].

The geography of the city under investigation must be reflected in a realistic manner within the computer in that as a patrol unit is assigned a call, we want to measure the distance and time traveled. In [10] and [36], the geography is organized by small polygonal shapes called atoms, the sum total of which form the city. Patrol beats and police districts are formed by combining these atoms. Crime statistics are assumed uniform across the area of an atom. In [38], the city is represented by a set of city block centroids and their coordinates (over 4,000 for Washington, D. C.) and the beats and districts are represented by a list of corresponding centroids. This model views the city as a discrete collection of points and all incidents which occur on a block are assumed to be located at the centroid.

The simulations are event-paced models. That is, once a certain set of operations associated with one event is completed, the program determines the next event that occurs and updates a simulation clock by adding to the present time the time until the next event. The program then proceeds with the set of operations associated with that event. Once the clock reaches some maximum time, the simulation is terminated and summary statistics are tabulated and printed out.

The main type of event that occurs is a reported incident or a call for police service. The time of occurrence of calls are generated by a Poisson process. The location of the call is determined from historical patterns which indicate the fraction of calls that originate from each atom or centroid. The priority of the call is determined from historical data which may vary by location.

Once the position and priority of an incident are known, the models execute the logic which attempts to assign a patrol unit to the incident. This algorithm is governed by the dispatch policy specified by the user. One component of the dispatch policy specifies the geographical area from which a unit may be dispatched. For example, only assign a unit whose patrol sector includes the geographical area containing the incident, or only assign a unit whose district designation is the same as that of the incident.

Given that a patrol unit is within the correct geographical area for a particular incident, the model then determines whether the unit is considered eligible for dispatch to this incident. This determination focuses on estimated travel time to the incident, the priority of the incident, and the current activity of the patrol unit. In general, the user may specify a dispatch policy that allows very important incidents to preempt (interrupt) patrol units servicing incidents of lesser importance. In addition, the importance of preventive patrol may vary with each unit, thereby giving the user the capability of assuring at least some minimal level of continuous preventive patrol. If no unit is found eligible for dispatch, the reported incident is inserted at the end of a queue of other unserviced incidents. There may be separate queues for each district and each priority level. If at least one unit satisfies the eligibility conditions, one is selected for dispatch according to a prespecified criterion such as minimal expected travel time. The assigned unit's priority status and position are changed accordingly.

A second major type of event occurs when a patrol unit completes servicing an incident. The logic is then executed that either reassigns the returning unit to an unserviced incident, or returns the unit to preventive patrol. The eligibility conditions regarding priorities, travel distances, and geographical areas, which are necessary to specify a dispatch policy, are also an integral part of the reassignment policy. In addition, it is necessary to specify how one unserviced incident is given preference over another [36].

The general structure of the simulation process is as follows. Inputs and initial conditions are classified into three groups: Geography, Parameters and Policies. Using this data, beats, patrol assignments, the simulation clock, dispatch queues, etc. are structured during initialization. Calls for Service (CFS) are generated to follow historical or hypothetical patterns. On the basis of dispatch policies, units are assigned to the CFS or are placed in queue. If a unit is assigned, travel, service and report/arrest times are generated using the input parameters.

Upon completion of service, a unit may be assigned to a CFS waiting in queue or it may resume preventive patrol.

During processing, operating statistics are gathered and, at the user's option, may be printed out. In addition, provision has been made to display a snapshot of the entire state of the system or to trace a number of units during the simulation. At the conclusion of the model run, summaries of the simulation performance are printed [38].

The input data requirements for simulation models are quite enormous and detailed. The geography data includes the structure of the city, districts, and beats (by atom or centroid). The operational parameters include the number of patrol units, types and their organization and beat assignment; times, locations, types and priorities of calls based on analyses of historical data; response parameters dealing with communication delay times, travel times, service times, arrest and report times. The gathering of the basic data for the inputs and initialization of a simulation model is an expensive and time consuming task. Most police departments are not now in a position to supply such data. Hopefully, as better data gathering procedures are initiated and computers introduced, the necessary parameters would be available.

In running a simulation to test out a new dispatch policy or beat design structure, there is no single measure of effectiveness which dominates all the others. In general, average response time, which can be measured by the simulation models, tends to be the overriding measure. However, the designers of the simulation models have allowed a wide range of measures to be gathered and presented to the decision-maker. These measures, plus the intuition and experience of the police commanders must be considered in making the final determination. There is a danger in relying on the computer selected solution—based on an agreed upon measure—to be field tested or implemented without considering possible impact of the solution in the total police decision-making framework.

The outputs and measures of effectiveness available from simulation models of the emergency response system include a wide range of statistical information which can be used to analyze the communications dispatch delays, travel time to incidents, service times, preventive patrol, cross beat and district service, workload of a patrol car, etc. A test computer printout of a simulation run of the Washington, D. C. simulation is shown in Figure 8. This output information can be presented to the police decision makers in many forms which will allow them to couple such data with their own experiences and intution. For example, simulation runs can be made which have all things constant except for a new district beat plan. The output measures for the new plan, especially beat workload and response time, can be compared to previous simulation runs and a decision made as to whether the new plan appears to be better. If so, then a field experiment with the new plan can be mounted with a higher level of success. In a similar fashion, we can evaluate major changes in the communications-dispatch area, e.g. computer aided dispatching such as New York City's SPRINT system, the advantages of a car locator system (see Reference [10] for a discussion of this type of system), the use of scooter patrols for dispatching to calls for service, and many other major as well as minor alternatives.

The simulation of the police emergency response system have taken two forms with respect to mode of operation. The Washington, D. C. system [38], which is in its final checkout stages at this writing, is run without any interaction from the analysts or users after they have supplied the input data; while the system reported in [36] and used for Boston can be run in an interactive fashion by the user who can input his data and assumptions via a

Simulation Run No. 35

Avg. Response Time
2.0 Min. for 17 CFS having priority 1
2.2 Min. for 19 CFS having priority 2
3.3 Min. for 14 CFS having priority 3

Time servicing CFS	30 percent
Time on preventive patrol	68 percent
Time on administrative out of service	0 percent
Time traveling to CFS	2 percent
Time for report & arrests	0 percent

	Overall	Prio. 1	Prio. 2	Prio. 3
CFS in waiting line				
Percent placed in waiting line	12.0	11.8	10.5	14.3
Avg. time (minutes)	0.1	0.1	0.1	0.2
Percent exceeding 360 sec.	0.0	0.0	0.0	0.0
Travel to incident				
Avg. time (minutes)	1.7	1.7	1.3	2.2
Percent exceeding 720 sec.	0.0	0.0	0.0	0.0
Avg. distance (miles)	0.4	0.4	0.3	0.5
Percent exceeding 1.0 miles	8.0	5.9	5.3	14.3
Response time				
Avg. time (minutes)	2.7	2.7	2.2	3.3
Percent exceeding 900 sec.	0.0	0.0	0.0	0.0
Service time				
Avg. time (minutes)	16.8	12.4	18.1	20.4
Report and arrest time				
Avg. time (minutes)				

FIGURE 8—*Computer Simulation Model of Police Dispatching and Patrol Functions*

terminal. Both systems have output procedures which yield different levels of detail, depending on the needs of the analysis and the user's experience.

One additional use for such simulation models is for the training of new personnel and for demonstrating to police personnel how their individual actions impact the operation of the total system. New dispatch personnel can be presented with a scenario of calls for service and asked to respond. The simulation can then show how their responses did in terms of selected output measures. District commanders can obtain more insight into the total response system by working with the simulation in a gaming exercise, where, for example, they are given a set of patrol units which they must deploy against a set of computer generated calls for service. Extreme situations, e.g. multiple bank holdups, can be put into the scenario. This type of training would probably be more effective if the simulation was tied to computer terminals so that the gaming and training could be run in real, elapsed time, e.g. an eight-hour shift over eight hours of simulated calls.

Perspective of Law Enforcement Models

From the above discussions we have seen that models of many types have been developed and are being used by law enforcement agencies. Much of the more advanced work has been supported by grant and other monies from Federal and State offices. To date there has been little carry-over of this work from the originating agency to other agencies. This has been due to lack of continuing funds, lack of technical personnel, and the inherent problems associated with documentation and understanding of technical computer-based models. There is no tradition for research and development within law enforcement agencies and professional personnel familiar with management science and operations research models are quite rare. It is hoped that the major police departments will fund such personnel, along with continuing Federal support. Until that time, the rich base of decision models for law enforcement will continue to be underutilized and underdeveloped.

(There are indications that the use of computer-based law enforcement information systems have been and will continue to be a prime mover in increasing the technical abilities of an agency and its use of models for decision making. Reference [39] shows that every State and over 100 local jurisdictions have such computerized systems, some of which incorporate decision-making models for the various components of the LE/CJ.)

III. COURT MODELS

As noted earlier, court system decision models have been directed mainly at court management problems. In what follows, we shall discuss some of the more viable approaches to these problems. There are related problem areas which confront the prosecuting attorneys, public prosecuters and defending attorneys. Little modeling work has been directed in these areas, but we note that some of the concepts discussed below, e.g. calendaring, simulation, should be of value to this side of the court system.

Court Statistics

In attempting to apply any type of analysis to the management of a court system, we are immediately faced with two interrelated problems: what measures do we use to evaluate possible changes in court operations and what data can be obtained to aid in such evaluations. An approach to what data and how such data should be gathered in order to modernize court administration is addressed in [11].

To illustrate how appropriate data can be used to compare court systems, we cite an experiment in the use of court statistics as reported in [40]. That study was directed towards developing measures for courts in large urban jurisdictions in order to provide a quantitative basis for making comparisons among cities as to how effectively their local court systems were operating. The selected indicators or measures of effectiveness considered were the amount of time taken to dispose of criminal cases; the extent to which those convicted have entered pleas of guilty; the percentage of jail prisoners awaiting trial; the amount of time prisoners spend awaiting trial; the backlog of criminal cases relative to the court's caseload; the average number of cases disposed of per judge; and the extent to which probation is used as an alternative to imprisonment.

These seven indicators do not exhaust the possibilities. Many more could be used to measure other aspects of the judicial process. A more complete list could include the following: the number of trial days; the length of prison sentences according to type of crime; cost per trial; cost per judgeship; the extent to which convictions are appealed; the portion of convictions overturned; the portion of defendants having adequate counsel, and so on to cover more detailed aspects of judicial processes. The selection made for this study was determined essentially by the importance of the measure and by the availability of source data for its calculation. The lack of suitable data has severely limited the choice of both indicators and cities.

As noted in [40], with few exceptions, record keeping in court systems is in a primitive stage; the most rudimentary management information needs are not being met in most jurisdictions. Court statistics that are available are fragmentary and in many instances the figures that appear in court documents are poorly defined.

The results of the survey and data collection effort are gathered in Figure 9. Some cities were omitted from the summary, primarily in instances where only a single indicator was available. For certain indicators where measures were available for only a few cities, all of the cities were included. Even though some selectivity was involved in terms of the cities included in the summary,

Summary

Court Indicators for Urban Areas

Urban Area	Number of Days from Indictment to Trial	% Convicted of Pleas of Guilty	% of Jail Prisoners Awaiting Trial	Ratio of Cases Pending to Dispositions	Number of Cases Per Judge	% Placed on Probation
Los Angeles	115	60	55	20	—	91
San Diego	101	89	45	.14	—	88
San Francisco	161	96 •	55	18	—	90
Washington, D.C.	285	—	33	.17	—	—
Urban New Jersey	142	—	—	—	—	—
Philadelphia	125	67	73	—	—	63
Chicago	—	83	65	—	—	34
Pittsburgh	—	53	38	37	1261	—
Detroit	—	86	52	31	—	66
Minneapolis	—	77	33	15	331	—
Dallas	—	92	75	—	—	36
St. Louis	—	77	49	—	—	—
Miami	—	62	—	—	—	—
Baltimore	174	—	76	—	655	—
New York	—	97	48	20	—	—
MEAN	158	78	54	21	749	66

FIGURE 9

Source: Reference [40]

the fragmentary nature of the data becomes obvious from the missing data points.

The purpose of the summary is to assess particular court systems on the basis of a number of indicators. It can be seen from the summary that the courts in some cities are performing better than average according to certain indicators and poorer than average according to others. This lack of agreement is demonstrative of the risk involved in assessing courts on the basis of a single indicator. The use of indicators in combination does not eliminate the need for interpretation of the data but does give a picture of general court performance.

The analysis illustrated above is an example of a rare, but important study of court operations. We still do not know, however, what values of the indicators show that a court system is working well. As we make changes in the court operations, e.g. more judges, a time history of the indicators would at least let us know if the change was in the right direction.

Court Planning and Operations

Research activities have produced a number of models which can be applied to decisions relative to court planning and operations. We do not believe that any of these models have gone beyond the research stage. Whether this is due to failings of the models or to the inability of court administrations to absorb and use new management procedures is, at this time, unclear. There are certainly few people who have intimate knowledge of court operations and problems who can comprehend and judge the adequacy of the research work. There are fewer who can implement these type of models. Like in law enforcement, funds and a commitment to investigate such advances for court systems are lacking. Thus, in what follows, we shall describe some models which appear to be of value in analyzing specific court problems.

The study [12] describes particular analytical models applicable to the planning and scheduling problems of a large criminal court system. A wide range of measures of effectiveness can be employed: minimize use of critical resources subject to acceptable processing times, maximize the number of hearings processed by the court, minimize the average waiting time for all cases, [12]. No single measure applies across the range of problems, and thus, the analyst and court administrator must augment the results obtained by a model with the usual experiential and intuitive factors associated with good management.

In evaluating alternative ways of operating a system, we must develop a proper cost function which expresses the relationships between the applicable operating costs and decision variables. For a court system we have cost of waiting (for defendants, lawyers and witnesses involved in the case), and the cost of overscheduling the system (calling more cases than can be heard in a single calling period) and underscheduling the system (running out of cases to hear before the end of a calling period) [12].

A queuing model approach to a court system treats the hearings to be processed as the arrivals and the judges as servers. We can assume that the arrival rate and service time distributions are Poisson and exponential, respectively, and look at a court system as a multi-server random process. Here it might be appropriate to separate the set of cases which arrive by type of priority. Standard formulas then allow us to determine average waiting time and average length of a queue, given the number of servers; or given the desired operating rates of a multi-server system, we can determine the minimum number of judges required to keep the system in equilibrium. Such models are discussed in [12] and data, which had to be gathered especially for the study, were used to test the model against present operations in Allegheny County, New York.

If we look at a court system in terms of the allocation of given resources, we can formulate certain decision problems as linear-programming models. In particular we consider a model which maximizes the number of hearings processed, where the decision variables are the number of hearings of each type to be scheduled in a time period [12]. This type of model can be used to evaluate planning changes in the court system or to see how a system will behave based on future projections on the number of hearings which will occur. Thus, bottleneck situations in future time periods or oversupply of judges can be anticipated. For this model we define

s_t = the number of judge-hours in period t, $t = 1, \ldots, n$,

h_j = the number of hearings of class j, $j = 1, \ldots, r$, arriving each period. Arrivals in period t cannot be processed until period $t + 1$.

p_j = the number of judge-hours required to process a hearing of class j.

b_j = the backlog of hearings in class j at the beginning of the periods.

x_{jt} = the number of hearings of class j scheduled in period t.

The constraints due to the number of available judges are given by

$$\sum_{j=1}^{r} p_j x_{jt} \leq s_t \qquad t = 1, \ldots, n$$

The constraints due to the need to process the backlog are given by

$$\sum_{t=1}^{k} x_{jt} \leq b_j + (k-1)h_j \qquad \begin{array}{l} j = 1, \ldots, r \\ k = 1, \ldots, n \end{array}$$

and we require

x_{jt} = nonnegative integers.

The linear-programming model is to maximize $\sum_{j=1}^{r} \sum_{t=1}^{n} x_{jt}$ subject to the above set of constraints. If we solve this problem by continuous linear-programming procedures the x_{jt} will not, in general, be integer valued; but a simple rounding scheme could suffice. This model is rather basic in its form and solution (schedule the shortest hearings first), but it can be strengthened if additional constraints are added which induce more realism, e.g. force a minimum of hearings of each case to be heard, insure against hearings being delayed too long. Similar models can be developed which minimize the number of judge hours per period and minimize the total waiting time of all hearings.

Another approach to analyzing court changes on delay in a criminal court is given in the report [41]. Here standard statistical procedures, e.g. regression analysis, were used to evaluate three possible alternatives to speeding up court operations: increasing the number of criminal case judges, increasing the judges' time on the bench, and a reduction in criminal jury trial time. Data from the Criminal Division of the U. S. District Court for the District of Columbia were used. A linear regression equation was developed which related, for each quarter, the number of workdays having the same number of trial starts to the number of trials commencing on a workday, i.e. a function which described the probability that there would be x trials commencing on any given workday in a particular quarter. For the data studied this turned out to be a linear function. This function was used to determine new functions for each of the proposed alternatives. For these alternatives, it turned out that none of the courses of action individually solved the problem of court delay. The course of action of adding one more judge reduced the growth of court delay to the slowest rate of growth. The addition of one judge full time as well as a second judge for approximately one quarter would stop the delay from increasing. Of course these conclusions assumed that the caseload of the District Court would not increase.

The possibilities were then investigated as to what would happen to delay if combinations of courses of action were implemented. In each instance assuming no increase in the case load, the expected backlog and expected waiting time for the cases were reduced. In the instance of implementation of all three courses of action, there was a 28.3% decrease in case backlog and a 39% decrease in expected waiting time over a period of one court year. This indicated that the best approach would be the implementation of several courses of action rather than relying on one method.

As noted in [11], "of all the court's administrative tasks the one most difficult is to schedule its proceedings. Schedules setting specific trials for a fixed day are next to impossible to achieve on either the criminal or the civil side. Schedules must nearly always be tentative. This may be the most trying weakness of judicial administration and is caused by a combination of the workload problems, administrative methods, and the lengths to which courts are willing to go to accommodate the convenience of attorneys and parties.

"The term 'calendar' seems to be used in a number of different senses, sometimes to mean all cases ready for trial, but more often in the narrower sense of only those which clerks have tentatively earmarked for trial during the court's current calendar year or session. The calendar may be viewed as those cases that have been put into waiting order for trial." Models and computer-assisted procedures for aiding this process have been investigated for a few court systems, [42] and [43]. The need is manyfold: a scheduled case is held over due to overloading, causing losses in time and possibly money by witnesses, lawyers and policemen; or not enough cases have been scheduled and prepared and a judge has to close down early. A specific approach to this problem was developed for the New York City Criminal Court, but it was not implemented as the Court's present organization structure and data processing abilities did not match what appear to be inherent requirements of such models.

The major objectives or measures of effectiveness of this model were to prepare schedules so that: the probability of a judge running out of cases in a given day is low; the probability that a case would be adjourned due to calendar overload is a minimum; attorney and policeman conflicts are eliminated or minimized; policemen's court appearances are batched and scheduled for duty days; high priority cases are scheduled as early as possible; a maximum time is allowed prior to which a case must be calendared. In addition, the system would free the judges and clerks from the chore of preparing and calling calendars, and date assignments would be made in real time so that participants in a case are informed immediately as to the time of their next court appearance and concurrence is obtained on the assigned data [42].

The model formulation is based on a set of logical rules dealing with priorities of the cases in backlog, available judicial time, and date preferences of the participants. The logical model was tested out for a simulated court system and did appear to perform properly. However, it was felt that basic changes in the administration and operation of court systems are a necessary prerequisite to the proper solution of the scheduling and other court problems. We also add that such changes should include the employment of technical and other specialists, and a continuing commitment to fund research and development efforts to study specific court operational problems.

Associated with the case scheduling problem is the problem of juror selection and juror waiting time reduction. A study conducted in the U. S. District Court for the District of Columbia covered juror utilization on 245 cases [44]. It showed that an average of 125 jurors were used each day from an available pool of 250. By use of a simulation model, it was found that in order to have a jury pool sufficient to meet expected daily peak demand for jurors, a 35 percent underutilization of the pool would be necessary (compared to the actual 50 percent). Although the approach in [44] was not to develop a particular decision model, it did show that the jury utilization process does have nice statistical measures and properties which could be used for a more detailed decision model. Necessary measures of effectiveness and trade-off considerations, e.g. cost of not having a trial due to jury shortage versus jury costs and waiting time, must be formulated and related to the problem of determining the correct size of the jury pool.

Court Simulation Models

As the above court models do not reflect fully the complex interrelationships of a court system, a more general approach to the understanding of a specific court system is by the development of a computer simulation model. Such simulations enable us to measure experimental changes on the system prior to their possible introduction into the ongoing system. These type of experiments on operational systems are somewhat difficult, costly and not fully controllable.

The President's Crime Commission in its Science and Technology Task Force Report [9] described a simulation model for the processing of felony defendants in the District of Columbia trial court system (see also [45]). The main emphasis of this study was to investigate proposed changes directed towards reducing delays in the court process. As hardly any two court systems are similar, a court simulation can not be devised for a general case, but must be directed towards a specific court environment. The Task Force study [45] investigated the following structure of the District of Columbia trial court system, as shown in Figure 10.

The system is divided into three parts: (1) the Court of General Sessions (the municipal court of the District of Columbia) and the U. S. Commissioner where defendants are brought for preliminary processing, (2) the grand jury section for indictments or informations, and (3) the U. S. District Court for processing from arraignment to the final disposition. The processing of felony cases was the only part of the court system examined in detail because the only adequate data available were concerned with this part. The numbers in Figure 10 indicate the flow of people accused of felonies through the system in 1965.

Figure 11 shows some of the major

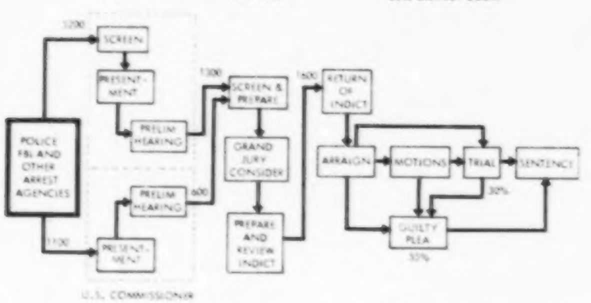

COURT OF GENERAL SESSIONS
GRAND JURY
U. S. DISTRICT COURT
1200
SCREEN
PRESENT-MENT
PRELIM HEARING
1300
SCREEN & PREPARE
1600
RETURN OF INDICT
POLICE FBI AND OTHER ARREST AGENCIES
GRAND JURY CONSIDER
ARRAIGN
MOTIONS
TRIAL
SENTENCE
PRELIM HEARING
600
30%
PREPARE AND REVIEW INDICT
GUILTY PLEA
55%
1100
PRESENT-MENT
U. S. COMMISSIONER

FIGURE 10—*Processing of felony defendants in District of Columbia courts (1965 data).* Source: Reference [45]

steps in the processing of felony defendants in the District of Columbia and the measures of time between steps obtained from the 1965 data, e.g. the median elapsed time between the data of preliminary hearing and return of indictment was 33 days, and the 80th percentile time was 54 days.

The simulation was calibrated and validated for the 1965 data and showed that a queue had developed in the grand jury unit. This was a queue of cases waiting to be processed through the grand jury unit either prior to or after grand jury consideration. Typically, in the simulation, a defendant's case spent 36 days in this queue waiting to be processed. An additional grand jury was simulated and added for one-third of the time, and an Assistant U. S. Attorney and a clerk were added for the full time. The results of the experiment showed that the time in queue was reduced from an average of 36 to 8 days with no significant increase in time after arraignment. The time from presentment to arraignment was reduced to 25 days as opposed to the 53 days observed in the 1965 data.

The problem of reducing the total time with emphasis on the time from arraignment to end of motions was studied by imposing certain procedural changes, including reduction of the guilty plea rate. It was found that the time for arraignment to first day of trial was about half of that observed, namely 50 instead of 100 days. There was, however, a longer time from end of motions to trial. This is a result of introducing less guilty pleas, i.e., more defendants go to trial. A queue waiting to be tried had developed.

Other simulation runs investigated

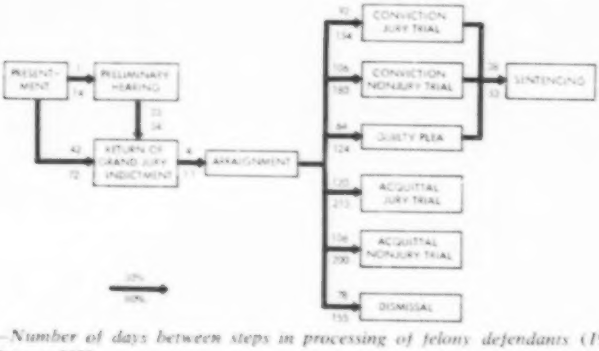

PRESENT-MENT
1
14
PRELIMINARY HEARING
33
54
42
72
RETURN OF GRAND JURY INDICTMENT
4
11
ARRAIGNMENT
92
134
CONVICTION JURY TRIAL
106
180
CONVICTION NONJURY TRIAL
26
53
SENTENCING
64
124
GUILTY PLEA
120
213
ACQUITTAL JURY TRIAL
106
200
ACQUITTAL NONJURY TRIAL
78
155
DISMISSAL
50%
80%

FIGURE 11—*Number of days between steps in processing of felony defendants (1965 data).* Source: Reference [45]

the effects of further increasing the grand jury resources and testing variations of procedural rules. These analyses showed that, if the input data represented the 1965 court system, a significant reduction in time could have been achieved, i.e., from an average of five to six months to an average of two to three months. Other simulation runs addressed the question of resource utilization.

As noted in [45], simulation can aid the courts in planning, programming and budgeting for future personnel and facility requirements, and for evaluating alternative resource allocations and operating procedures. It can be used by the court administrator to investigate and compare contemplated but untried operating concepts, thereby allowing him to experiment with possible changes before putting them into effect. Finally, it can perhaps lead to a fuller understanding of the entire court system—how it operates and how various elements interact.

We must also stress here, as was done for most of the models developed for the LE/CJ area, that simulation models require a detailed logical structure of the system and a firm set of historical data. It is rare to find such elements or even means for developing them, e.g. data standards and definitions, within court systems. Simulation models for courts, as well as the other court models have made little impact on the planning and operational aspects of our courts. Here we do not need new technical advances, the state-of-the-art is more than adequate. As in law enforcement, the court systems need a continuing commitment to improvement which calls for funds, legislative and procedural changes, and the introduction of qualified technical professionals as part of the court system.

IV. CORRECTIONS SYSTEM MODELS

There is a paucity of models and research into decision making as applied to correctional problems. The deep behavioral aspects associated with these problems does, of course, hinder logical and mathematical approaches. However, statistical techniques as aids to sentencing and selection of proper treatment of individuals, and as a means of analyzing different correctional programs appear to be feasible [9]. Research in these areas is still embryonic, with references [46], [47], [48] and [49] as typical efforts. We do not know of any specific models which have been employed in an operational context. The study [48] made for the California State assembly was able to show, for the California prisons, that recidivism was not found to be significantly related to time-served, but is closely correlated to the characteristics and history of the individual. These results did influence new parole procedures and parole decision guidelines.

One of the key problems of rehabilitation is that of selecting a particular rehabilitative program to implement from among several alternatives. An approach to this decision-making area was developed by the Science and Technology Task Force [9]. This model was based on analyzing the average number of career arrests and the costs of those criminal careers under different circumstances: the current rehabilitative program and new proposed programs. Changes in the career arrest rate and costs due to these new programs were assumed (although they could and should be obtained experimentally). The model established tradeoffs in terms of how much money could be spent on programs that reduce rearrest probabilities and thus, total criminal career costs. This model approach, if it was combined with a serious effort to obtain data based on actual controlled experimentations, could be used to evaluate such programs in terms of recidivism rate and correctional system costs, thus allowing for more rational decision-making framework to be introduced into this critical field.

V. HOLISTIC APPROACHES TO ANALYSIS OF THE LE/CJ SYSTEM

The above descriptions of the decision-making problems and models have viewed the LE/CJ system in terms of its major component parts—police, courts and corrections. Solutions afforded by these models tend to work in isolation and neglect the secondary but important impacts which a change in one component can cause in the others, e.g. the situation which could arise if a new approach was taken to the arrest and processing of drunks. In general, little consideration has been given in the past to these interrelationships, especially in terms of planning. However, with the increase of Federal funds to States and local agencies, criminal justice planning agencies now exist in most major jurisdictions, and some of them have demonstrated that integrated planning is feasible.

The Safe Streets Act of 1968 enables each State to establish and operate a State criminal justice agency, supported by funds from the Law Enforcement Assistance Administration of the Department of Justice. Representation on an agency must include a supervisory group that represents all elements of the LE/CJ system, as well as citizen groups and local government units. Each agency must develop and submit a comprehensive State plan. Although the implementation of general revenue sharing may change a State's approach to the development of such a plan, the following comments are appropriate.

Each plan focuses on the problem of crime: how much there is, what causes it, how to prevent it, how to control it, how to treat people who commit crimes, and how to improve and expedite justice. It examines the physical and human factors that produce crime, and analyzes the needs of police, prosecutors, defense attorneys, courts, the correctional processes and the offenders. Each State plan must offer realistic, specific goals; be action-oriented; and weigh costs and benefits.

The State plan sets overall priorities for each major LE/CJ area and priorities for goals at the State and local levels. It incorporates innovations and advanced techniques; describes general needs and problems, existing systems, available resources, and administrative machinery for implementing the plan. It defines the direction, scope and general types of improvements to be made, and also encourages units of local government to make cooperative arrangements with respect to services, facilities and equipment [50].

To aid in analyzing such comprehensive plans, agencies have been able to use cost-benefit, program budgeting and similar planning techniques. A holistic LE/CJ model for evaluation of alternative plans in terms of their impacts across the system was developed by the Science and Technology Task Force [9], [51] and [52]. (Other holistic models are described in [53], [54] and [69].)

This model considers the flow of offenders through the many branches and stages of the LE/CJ system (Figure 1) and based on costs and probabilities associated with passing through a branch or stage, determines the workloads and total costs under varying assumptions of how the components do or should operate. The following description is excerpted from [51] where two approaches to this type of model are developed. First, there is a simple production process, in which the principal concern is the flow through the system, and the accumulation of costs flowing from a single arrest. This linear model provides an opportunity to examine at each stage the workload, the personnel requirements that result, and the associated costs; to attribute these to types of crimes; and to project all of these planning variables as functions of future arrest rates. The second is a feedback model, which considers the recidivism probability associated with each released defendant, and his subsequent processing for future arrests after he has once been released by the LE/JC system. Such a feedback model permits estimating the costs of a total criminal career and the consequences of alternative actions within the LE/CJ system to lower recidivism probabilities.

Briefly, the linear steady-state model is as follows [51]. It is used to compute the costs and workloads at the various processing stages and to establish manpower requirements to meet the anticipated workloads. The workload is the annual demand for service at the various processing stages (e.g., courtroom hours, detective man-hours); the manpower requirement is derived from the workload by dividing the annual working time per man; total operating costs are allocated to offenders by standard cost-accounting procedures.

The flow of persons through each processing stage is described by a vector whose i^{th} component represents the yearly flow associated with characteristic type i ($i = 1, \ldots, I$). These characteristics can be any attribute associated with individual offenders, their crimes, or their previous processing by the LE/CJ system.

The independent flow vector to the model, which must be specified as input, is the number of crimes reported to police during one year. The outputs are the computed flows, costs, and manpower requirements that would result if the input and the system were in steady state. Each processing stage is characterized by vector cost rates and branching probabilities.

It is recognized that the model does have certain simplifying features and assumptions, but, given a set of consistent data, it does yield acceptable first estimate costs, workloads and flows by crime type and processing stage. More important, the model measures the effects of changes in one component on the workload, cost and manpower requirements of the other components, e.g. changes in crime or arrest rate, probation procedures, and also enables the determination of future resource requirements based on projections of expected crime level.

The second model, the feedback model, describes the recycling process through the LE/CJ systems during the course of an individual's career. The applications of this model are far-reaching for planning purposes: given the age of an offender at first arrest and the crime for which he is arrested, the model computes his expected criminal career profile (i.e., the expected crimes for which he will be arrested at each age); using the cost results of the linear model, the model computes the average costs incurred by the LE/CJ system over a criminal career; recidivism parameters (e.g. rearrest probabilities) can be varied to assess how each parameter affects criminal careers and cost; and the model provides a unified framework in which to study the process of recidivism and in which to test the effects of proposed alternative LE/CJ system policies on recidivism.

The feedback model structure is as follows [51]. As in the linear model, flows are distinguished by crime type. In addition, each flow variable is broken down by the offender's age. The input to the model, rather than crimes reported to police, is the number of arrests during a year, by crime type and by age, of individuals who have never previously been arrested for one of the crimes being considered. In the model, these "virgin" arrests are added to recidivist arrests to obtain the total arrests during the year. The total arrests then proceed through the LE/CJ system just as they do in the linear model.

Since the offender flows comprise individuals who cycle back into the system after dismissal or release from the LE/CJ system, it is necessary to compute the number that do recycle when they are rearrested and for what crime. At each possible dismissal point, the offender is characterized by a probability of rearrest that is, in general, a function of his age and his prior criminal record. The expected number who will be rearrested at some later time is computed by multiplying the number in the flow by the appropriate rearrest probability. Then, the age at rearrest is computed by using the distribution of delay between release and the next arrest. Finally, the crime type of the next arrest is computed from a rearrest crime-switch matrix, where the matrix element is the conditional probability that the next arrest is for crime type j, given that rearrest occurs and the previous arrest was for crime type i.

The data requirements for the above models are quite demanding and few, if any, jurisdictions are in a position to supply the requisite data or to conduct the necessary studies and data collection procedures. The models were originally tested with composite data which originated from California and other jurisdictions. More recently, the linear model has been modified and applied to the City of Philadelphia and the State of Alaska with good results [55]. The actual Philadelphia model has 28 processing stages, 89 branches, 89 manpower types, 20 non-salary cost types and 29 crime types. Alternative plans are evaluated against the base LE/CJ structure and data and comparisons can be made on such indicators as court backlog, probation officer caseload, and average population of prisoners.

These models can also play an important role in demonstrating to planners how their actions effect the other elements of the LE/CJ system. To this end, the linear model has been programmed to be run on a computer in an interactive terminal mode and has been used as a training device [52], [56]. This program is used principally as a design tool for testing a variety of changes in the operation of the LE/CJ system. These changes might include introduction of additional public defenders, discontinuation of arrests for certain types of crimes, introduction of additional judges or introduction of innovative correctional programs.

In many cases, there is no theoretical or empirical basis for assessing the detailed consequences of a contemplated change. In view of this difficulty, the appropriate persons to make these estimates are the LE/CJ system planners and they can do this with an interactive model. In response to inquiries from the program, the criminal justice planner at a terminal makes changes to the system parameters, runs the model, and then notes the resource implications of those changes. Thus, he can quickly try a variety of changes, evaluate their consequences, and use that feedback of consequences to try further modifications.

The need for holistic models of specific LE/CJ systems is apparent. As evidenced by the above models, we do have valid approaches to the study of the LE/CJ system, given that assumptions and refinements must be tailored to each system. But these models cannot be made functional unless concerted efforts are made throughout all elements of the system to mount continuing data gathering activities. Only in this manner will we ever be able to develop measures and perform evaluations of the LE/CJ system.

References

[1] Mode, E. B., "Probability and Criminalistics," *American Statistical Association Journal*, September 1963, pp. 628–640.

[2] *The Challenge of Crime in a Free Society*, The President's Commission on Law Enforcement and Administration of Justice, U.S. Government Printing Office, Washington, D.C., February 1967.

[3] *State-Local Relations in the Criminal Justice System*, Advisory Commission on Intergovernmental Relations, U.S. Government Printing Office, Washington, D.C., 1971.

[4] *Criminal Justice Agencies in the United States Summary Report 1970*, Department of Justice, U.S. Government Printing Office, Washington, D.C., 1970.

[5] "Report of the National Advisory Commission on Criminal Justice Standards and Goals," Government Printing Office, Washington, D.C., 1973.

[6] *Crime in the United States*, Uniform Crime Reports, FBI, U.S. Government Printing Office, Washington, D.C., 1971.

[7] Kakalik, J. S. and S. Wildhorn, "Aids to Decision-making in Police Patrol," (R-593-HUD/RC) The Rand Corporation, Santa Monica, California, February 1971.

[8] Gallup Poll: "One in Three is City Crime Victim" and "Crime is Rated Worst Urban Problem," The Washington Post, January 14 and 15, 1973.

[9] *Science and Technology: Task Force Report*, The President's Commission

on Law Enforcement and Administration of Justice, U. S. Government Printing Office, Washington, D. C., 1967.

[10] Larson, R. C., *Urban Police Patrol Analysis*, The MIT Press, Cambridge, Massachusetts, 1972.

[11] *The Courts: Task Force Report*, The President's Commission on Law Enforcement and Administration of Justice, U. S. Government Printing Office, Washington, D. C., 1967.

[12] Lyons, N. R., "Analytic Models of Criminal Court Operations," Graduate School of Business and Public Administration, Cornell University, Ithaca, New York, 1971.

[13] Corrections Task Force Report, The President's Commission on Law Enforcement and Administration of Justice, U. S. Government Printing Office, Washington, D. C., 1967.

[14] Pittman, J. T. and P. Gray, "Evaluation of Prison Systems," School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia, November 1972.

[15] Roy, R. H., "An Outline for Research in Penology," *Operations Research*, January–February, 1964.

[16] "Resource Allocation at the St. Louis Metropolitan Police Department," IBM Application Brief, K20–0230–0, IBM, White Plains, New York.

[17] *Annual Report, Fiscal Year 1970*, Metropolitan Police Department, Washington, D.C.

[18] Smith, R. D., "Computer Applications in Police Manpower Distribution," International Association of Chiefs of Police, Gaithersburg, Maryland, 1964.

[19] McEwen, J. T., "A Mathematical Model for Prediction of Police Patrol Workload," St. Louis Police Department, St. Louis, Missouri, 1966.

[20] Crowther, R. F., "The Use of A Computer System for Police Manpower Allocation in St. Louis, Missouri," Department of Police Administration, Indiana University, Bloomington, Indiana, 1964.

[21] Shumate, R. P. and R. F. Crowther, "Quantitative Methods for Optimizing the Allocation of Police Resources," *Journal of Criminal Law, Criminology and Police Science*, Vol. 57, pp. 197–206, June 1966.

[22] LEMRAS, IBM Program Product, Report No. H20–0629–0, IBM White Plains, New York.

[23] Brown, R. G., *Smoothing, Forecasting and Prediction of Discrete Time Series*, Prentice-Hall, New Jersey, 1962.

[24] Gass, S. I., "On the Division of Police Districts Into Patrol Beats," *Proceedings—1968 ACM National Conference*.

[25] Heller, N. B. et al., "Computer Design and Evaluation of Police Patrol Beats," St. Louis Police Department, St. Louis, Mo., 1970.

[26] Heller, N. B. et al., "Partitioning of Police Districts Into Optimal Patrol Beats Using a Political Redistricting Algorithm: Model Design and Evaluation," St. Louis Police Department, St. Louis, Missouri, 1971.

[27] Dean, B. V., et al., "A Preliminary Systems and Allocation Study of the Cleveland Police Department," Operations Research Department, Case Western Reserve University, Cleveland, Ohio, 1970.

[28] Hess, S. et al., "Nonpartisan Political Redistricting by Computer," *Operations Research*, Vol. 13, No. 6, 1965.

[29] "Crime Prediction Modeling," Jet Propulsion Laboratory, Task 125, Final Report, Pasadena, California, April 1971.

[30] Stein, D. P. et al., "Crime Prediction by Computer—Does It Work and Is It Useful," *Law Enforcement Science and Technology Vol. II*, S. I. Cohn, Ed., IIT Research Institute, Chicago, Illinois, 1969.

[31] Rosenshine, M., "Contributions to a Theory of Patrol Scheduling," *Operational Research Quarterly*, Vol. 21, pp. 99–106, 1970.

[32] Elliott, J. F., "Random Patrol," Technical Information Series, R67ELS–77, General Electric Corporation, Electronics Lab., Syracuse, New York, 1967.

[33] Larson, R. C., "Operational Study of the Police Response System," Technical Report 26, Operations Research Center, MIT, December, 1967.

[34] Bottoms, A., "Models for the Allocation of Urban Police Patrol Forces," Vols. I and II, Final Report, Grant No. 195, Office of Law Enforcement Assistance, U. S. Department of Justice, Washington, D. C., November 1969.

[35] Bennett, W. and J. R. DuBois, "The Use of Probability Theory in the Assignment of Police Patrol Areas," PR70–2, Law Enforcement Assistance Administration, Department of Justice, Washington, D.C., July 1970.

[36] Larson, R. C. and J. C. Williamson, "On-line Simulation of Urban Police Patrol and Dispatching," Urban Sciences, Inc., Wellesley, Massachusetts, 1972.

[37] Adam, R. et al., "Simulation of Police Field Forces for Decision-Making in Resource Allocation," Engineering Economic Systems, Stanford University, Stanford, California, June 1969.

[38] Miller, H. F., Jr. and B. A. Knoppers, "A Computer Simulation of Police Dispatching and Patrol Functions," MATHEMATICA, Inc., Betheseda, Maryland, October 1972.

[39] "1972 Directory of Automated Criminal Justice Information Systems," U.S. Department of Justice, LEAA, December 1972.

[40] DonVito, P. A., "An Experiment in the Use of Court Statistics," *Judicature*, Vol. 56, No. 2, August–September 1972.

[41] Reed, J. H., "The Application of Operations Research to Delay in a Criminal Court of Original Jurisdiction," Division of Business Administration, Clarion State College, Clarion, Pennsylvania, 1971.

[42] Shapiro, S. S., "An Automated Court Scheduling System," Programming Methods, Inc., New York City, New York, October 1971.

[43] Isaacs, H. H., "A Study of Witness Utilization and Case Scheduling in the San Bernardino County Municipal Court," Isaacs Assoc., Los Angeles, California, October 1970.

[44] Pabst, W. R. Jr., "Juror Waiting Time Reduction," PB–201 412, National Technical Information Service, Springfield, Virginia, June 1971.

[45] Taylor, J. G. et al., "Simulation Applied to A Court System," IEEE Transactions on Systems Science and Cybernetics, Vol. SSC–4, No. 4, November 1968.

[46] "Prediction of Criminal Conduct and Preventive Confinements of Convicted Persons," Andrew von Hirsch, Buffalo Law Review, Spring 1972.

[47] Clapp, D. E., "Parole Viewed as a Decision Under Uncertainty," *The Journal of Industrial Engineering*, August 1966.

[48] Pittman, J. T. and P. Gray, "Evaluation of Prison Systems," School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia, 1972.

[49] Daetz, D. and S. E. Kolodney, "Results and Impact of an Analysis of Recidivism in California," Department of Industrial Engineering, Stanford University, Stanford, California, 1972.

[50] "A United Strategy for Crime Control," Law Enforcement Assistance Administration, U. S. Department of Justice, 1968.

[51] Blumstein, A. and R. Larson, "Models of a Total Criminal Justice System," *Operations Research*, Vol. 17, No. 2, March–April, 1969.

[52] Blumstein, A. and R. Larson, "Analysis of a Total Criminal Justice System," Chapter 16 of *Analysis of Public Systems*, A. W. Drake, R. L. Keeney and P. M. Morse, MIT Press, Cambridge, Massachusetts, 1972.

[53] "Prevention and Control of Crime and Delinquency," Youth and Adult Corrections Agency, State of California, 1965 (prepared by the Space-General Corporation, El Monte, California).

[54] Abraham, S. C., "Simulation Modelling in Criminal Justice: An Approach and Demonstration Using System Dynamics," Graduate School of Management, University of California, Los Angeles, California, July 1972.

[55] Renshaw, B. H. et al., "**PHILJIM**, The Philadelphia Justice Improvement Model," Government Studies and Systems, Inc., Philadelphia, Pennsylvania, October 1972.

[56] Belkin, J. A., et al., "**JUSSIM**: An Interactive Criminal Justice System Design Tool," Urban Systems Institute, Carnegie-Mellon University, 1971.

Selected Additional References

General

[57] "Combating Crime," *The Annals of the American Academy of Political and Social Science*, November 1967.

[58] "Crime Against Small Business," A Report of the Small Business Administration transmitted to the Select Committee on Small Business, U.S. Senate, Document No. 91–14, April 3, 1969.

[59] "Criminal Statistics," National Institute of Mental Health, Center for Studies of Crime and Delinquency, Rockville, Md., DHEW Publication No. (HSM) 72–9094.

[60] "Economic Crimes: Their Generation Deterrence and Control," Harold L. Votey, Jr., et. al., NTIS, Department of Commerce, PB 194 984, 31 December 1969.

[61] "Evaluation of Crime Control Programs," Michael D. Maltz, Law Enforcement Assistance Administration, U.S. Dept. of Justice, #2700–00163, April 1972.

[62] "Federal Law Enforcement and Criminal Justice Assistance Activities," Attorney General's First Annual Report for 1971, #2700–0160.

[63] *Foundations and Mathematical Models of Operations Research With Extensions to the Criminal Justice System*, Haig E. Bohigian, The Gazette Press, Inc., 1971.

[64] "The Impact of Organized Crime on an Inner City Community," Harold D. Lasswell, Jeremiah B. McKenna, The Policy Sciences Center, Inc., New York, New York.

[65] "Implementing Statewide Criminal Justice Statistics Systems—The Model and Implementation Environment," Project Search, Law Enforcement Assistance Administration, January 1972, Technical Report No. 4.

[66] "Mathematical Methods in Criminology," M. E. Wolfgang and H. A. Smith, *International Social Science Journal*, Vol. XVIII, No. 2, 1966.

[67] "Property Crime as an Economic Phenomenon," David L. Sjoquist, U.S. Department of Commerce, NTIS, December 1970, PB–203–144.

[68] "Regional Criminal Justice Planning—A Manual for Local Officials," Part

I, National Association of Counties Research Foundation," Washington, D.C., June 1971.

[69] *Ibid.*, Part II, November 1971.

Law Enforcement

[70] "Allocations of Resources in the Chicago Police Department," National Institute of Law Enforcement and Criminal Justice, U.S. Department of Justice, PB–212777, March 1972.

[71] "Analysis of Crime Prevention Strategies," Ronald L. Rardin and Paul Gray, 42nd National Meeting of the Operations Research Society of America, November 1972.

[72] "The Challenge of Productivity Diversity—Improving Local Government Productivity Measurement and Evaluation," Part III, *Measuring Police-Crime Control Productivity*, June 1972, for The National Commission on Productivity by The Urban Institute.

[73] "The Collection and Analysis of Auto Theft Data in Denver," July 1970–June 1971, Final Report, Industrial Economics Division, Denver Research Institute, University of Denver, Anita S. West, March 1971.

[74] "Computerized Allocation of Police Traffic Services," A Demonstration Study, NTIS, U.S. Department of Commerce, PB–213 025, March 1972.

[75] "Computer Assisted Instruction Program For Police Training," Richard W. Brightman, U.S. Department of Justice, National Institute of Law Enforcement and Criminal Justice, Pr 71–5, December 1971.

[76] "Crime-Correlated Areas," Frank S. Rudnick, University of Rhode Island, April 1972.

[77] *Crime and Information Theory*, M.A.P. Willmer, Edinburgh University Press, 1970.

[78] "An Economic Analysis of the Distribution of Police Forces," J. F. Giertz, U.S. Department of Commerce, NTIS, 1 April 1970.

[79] "Measuring Auto Theft and the Effectiveness of Auto Theft Control Programs," Robert Frese, Nelson B. Heller, for 38th National Meeting of the Operations Research Society of America, Sept., 1970.

[80] "Measuring the Response Patterns of New York City Police Patrol Cars," Richard C. Larson, The N.Y.C. Rand Institute, R–673-NYC/HUD, July 1971.

[81] "Measuring the Value of the Indianapolis Police Fleet Plan," Donald M. Fisk, The Urban Institute, Washington, D.C., Working Paper #108–70, 9/10/70.

[82] "Methods for Allocating Urban Emergency Units," J. M. Chaiken and R. C. Larson, The N.Y.C. Rand Institute, R–680-HUD/NSF, May 1971.

[83] "Minority Recruiting in the New York City Police Department—Part I. The Attraction of Candidates—Part II. The Retention of Candidates," I. C. Hunt, Jr.—Part I., B. Cohen—Part II., The N.Y.C. Rand Institute, R–702–NYC, May 1971.

[84] "Nonlethal Weapons for Law Enforcement—Research Needs and Priorities, Security Planning Corporation, Washington, D.C., March 1972.

[85] "On the Measurement of Information in the Field of Criminal Detection," M.A.P. Willmer, Operational Research Quarterly, Vol. 17, No. 4.

[86] "Patterns of Burglary," Harry A. Scarr, U.S. Department of Justice, National Institute of Law Enforcement and Criminal Justice, PR 72–10, February 1972.

[87] "Operations Research and Computers in Law Enforcement: Some Case Studies of American Police Experience," A. M. Bottoms, *Computers and Operations Research*, Vol. 1, pp. 49–65, 1974.

[88] "Police and Computer: Use, Acceptance, and Impact of Automation," Kent W. Colton, Joint Center for Urban Studies of MIT and Harvard University, Municipal Year Book, 1972.

[89] "Police Background Characteristics and Performance: Summary," Bernard Cohen and Jan Chaiken, For the National Institute of Law Enforcement and Criminal Justice, The N.Y.C. Rand Institute, R–999–DOJ (abridged), May 1972.

[90] "Police Training and Performance Study," U.S. Department of Justice, National Institute of Law Enforcement and Criminal Justice, PR 70–4, September 1970.

[91] "Some Determinants of Crime: An Econometric Analysis of Major and Minor Crimes Around Boston," George B. Weathersby, University of California, for the 38th National Meeting of the Operations Research Society of America, Sept. 1970.

[92] "Some Effects of an Increase in Police Manpower in the 20th Precinct of New York City," S. James Press, The N.Y.C. Rand Institute, R–704–NYC, October 1971.

[93] "Urban Crime and Urban Planning. A Pilot Study, Vol. I," Hubert G. Locke, U.S. Department of Commerce, NTIS, PB 195 655, July 1969.

[94] "The Use of Gaming Techniques in Police Research," A. G. McDonald, Scientific Adviser's Branch, London, England, SA/POL 12, April 1966.

[95] "Watts Riot Arrests,"—Los Angeles, August 1965, Bureau of Criminal Statistics, Department of Justice, California, June 30, 1966.

Courts

[96] "Compilation and Use of Criminal Court Data in Relation to Pre-Trial Release of Defendants: Pilot Study," J. W. Locke, R. Penn, J. Rick, E. Bunten, and G. Hare, U.S. Department of Commerce, National Bureau of Standards, NBS Tech. Note 535, August 1970.

[97] "Evaluation of the Manhattan Criminal Court's Master Calendar Project: Phase 1—February 1–June 30, 1971," John B. Jennings, The N.Y.C. Rand Institute, R-1013-NYC, January 1972.

[98] "The Flow of Arrested Adult Defendants Through the Manhattan Criminal Court in 1968 and 1969," John B. Jennings, The N.Y.C. Rand Institute, R-638-NYC, January 1971.

[99] "The Flow of Defendants Through the New York City Criminal Court in 1967," John B. Jennings, The N.Y.C. Rand Institute, RM-6364-NYC, September 1970.

[100] "Quantitative Models of Criminal Courts," John B. Jennings, The N.Y.C. Rand Institute, P-4641, May 1971.

Corrections

[101] "The Bronx Sentencing Project of the Vera Institute of Justice," J. B. Lieberman, S. A. Schaffer, J. M. Martin, U.S. Department of Justice, National Institute of Law Enforcement and Criminal Justice, PR 72-15, October 1972.

[102] "Parole Decision Making—The Utilization of Experience in Parole Decision Making: A Progress Report," D. M. Gottfredson, L. T. Wilkins, P. B. Hoffman, S. M. Singer, U.S. Department of Justice, National Institute of Law Enforcement and Criminal Justice, January 1973.

[103] "The Measurement and Prediction of Criminal Behavior and Recidivism," W. O. Jenkins et al., Draper Correctional Center, Elmore, Alabama, December 1972.

Chapter 9

Models in Educational Planning and Operations

By

Edmond H. Weiss

SUMMARY 279

I. EDUCATIONAL DECISIONS 281
  Introduction 281
  Decision Levels 282
  School District Level: Student Centered Decisions 283
  Longer Run Student Decisions 284
  School District Level: Curriculum Decisions 285
  School District Level: Campus Operating Decisions 285
  School District Level: District-Wide (Or University-Wide) Decisions (Short Run) 286
  District-Wide Decisions (Long Run) 287
  District-Wide (Major Decisions) 288
  Intermediate (County) Level Decisions 288
  State-Level Decisions 289
  Federal Decisions 291

II. ILLUSTRATIONS OF DECISION MODELS FOR EDUCATION 292
  Introduction 292
  A Theoretical Model for Optimizing the Economy and Education 292
  Forecasting Nationwide Undergraduate Enrollments and Federal Financial Aid Requirements 294
  A Model for Allocation of Vocational Education Funds 296
  A State-Level Input-Output Planning Model for Local Educational Agencies 297
  A Model for College Enrollment Forecasts 298
  A Model for University Resource Requirements Analysis 299
  A Model for University Facilities Analysis 300
  A Model for School District Resource Requirements Planning 303
  A Model for School Bus Scheduling 304
  A Model for Class Scheduling 307
  A Theoretical Model for Instructional Optimization 309

III. AN ASSESSMENT OF DECISION MODELS IN EDUCATION 311
  Introduction: Assessment Criteria 311
  Assessment 312
  The Future of Models 313

REFERENCES AND SELECTED BIBLIOGRAPHY 313

Preceding page blank

Models in Educational Planning and Operations

SUMMARY

American education is a complex of institutions, governments and systems, providing essential services to more than 50,000,000 children and adults. Substantial segments of this complex can be found at every level of government (from national to neighborhood) and in both the private and public sector.

Decision models have grown increasingly prominent in education management. Most of the systems and techniques have been developed to support decision problems which education has in common with other large organizations: staff assignment, vehicle deployment, construction scheduling, cost forecasting, and many others. In addition, several decision models have been advanced to solve decision problems peculiar to education, such as: relating learning to instructional techniques, satisfying the skilled manpower requirements of the economy, or achieving racial integration in school programs.

In this chapter, a selection from the virtually unlimited range of educational decision problems is presented. The decisions are organized according to a hierarchy of institutions—from the classroom, to the school, to the intermediate unit, to the state agency, to the Federal Government. Interestingly, the decision problems discussed (chosen because of their relative importance and amenability to model-assisted solutions) indicate that each level of the education complex has similar decision problems: most of the differences are matters of scope, timing, and level of aggregation. Among the most prevalent decision problems at all levels are: forecasting of demand for service; calculation of costs and resource requirements for programs; allocation of resources to produce optimum results; assignment and scheduling of manpower and material resources, to satisfy demand and achieve economies; projecting the operational consequences of alternative administrative policies.

The chapter begins with a description of the complex, its levels, and the main decision problems. This section is followed by a sampling of representative and instructive models for national, State, and local levels (see Summary Table).

It should be noted that the majority of decision models presented here are not directly related to the conduct of instruction. Administrative questions can usually be more readily answered in quantified terms than instructional problems. It is vastly easier to construct simulations of the service systems in education than it is to establish direct and mathematically precise relationships between resource utilization and student growth. Even so, the benefits of these models will ultimately be shared by students, as a result of four related consequences of their use:

- More efficiency in the non-instructional aspects of education, "freeing-up" instructional resources.
- More efficiency in the selection and use of instructional resources, so that there is a better match between apparent student "needs" and resources.
- More educational research, of higher direct utility to educational practitioners.
- More formal and explicit information about the operation of schools

Preceding page blank

EDUCATION

Models Discussed in Chapter

Summary Table

Model Decision Area	General Type	Important Characteristics
National Level:		
Macro-Economic Planning (OECD)	Simulation Linear Programming	Determines optimum mix of national investment in education and non-education sectors of the economy, to insure economic growth, provision of required manpower, and a continuing level of "non-material" educational benefits
Forecasting undergraduate enrollments and financial aid requirements	Probabilistic Flow Simulation; Deterministic Resource Requirements Simulation	Forecasts total undergraduate enrollments in the U.S., by socio-economic group; calculates the financial aid burden, under varying policies of student/government financial obligation
State Level:		
Allocation of Vocational Education Funds (Penna.)	Linear Programming Simulation	Determines the optimum mix of grant awards to vocational education programs in the state, so as to maximize the closeness between skilled graduates produced and labor market demands
Input-Output Evaluation of Local School Districts in a State (New York)	Multiple Regression	Evaluates the effectiveness of programs in various school by comparing observed results with results "expected" in the multiple regression equations for the district in question
Forecasting State-Wide College Enrollments (California)	Probabilistic Flow Simulation	Forecasts the enrollment in a state-wide system of colleges, at each year-level, as a function of current enrollments in high school and college and transitional probabilities
College/University Level:		
Resource Requirements Prediction (WICHE)	Deterministic Simulation	Calculates staff, facilities, equipment, and other resource requirements as a function of enrollment by program; relies on an "induced course load matrix" which converts program service demands to resource requirements
Facilities Analysis Model (CCHE)	Deterministic Simulation Optimization	Shows the alternative costs of various policies on the use of higher education facilities in a given institution; shows minimum cost as a result of trading-off operating costs vs. capital costs
School District Level:		
Resource Requirements Model (Trenton)	Deterministic Simulation	Forecasts the multi-year costs and staff requirements for a school district, based on current utilization, change in demand, inflation, etc; calculates consequences of current policies or simulates the consequences of alternative plans and policies

Model/Decision Area	General Type	Important Characteristics
Vehicle Scheduling (Tippecanoe)	Heuristic Simulation	Determines school vehicle schedules which minimize bus routes, minimize mileage, overt overloading; and keep route time below constraint ceilings
School Level: Class Scheduling	Monte Carlo Simulation	Assigns students to courses, minimizing the arbitrary effect of "closing" sections on a "first-come, first-served" basis; minimizes the number of "unassignable" cases
Classroom Level: Optimizing Learning	Linear Programming	Develops a theoretical model for assigning students to tasks, so as to maximize achievement of desired performance objectives

and colleges, so that "institutional learning" will take place, and all decisions will be more enlightened.

The chapter concludes with an assessment of the adequacy of current models (and forecasts of needed and probable changes), as well as a resource bibliography which will aid education administrators to apply models to their problems.

I. EDUCATIONAL DECISIONS

Introduction

The purpose of this chapter is to introduce educational policymakers, administrators, and program managers to the range of decision problems for which decision models are and (in some cases) are not available, and to give general and specific examples of the models themselves. No technique or model is described here in sufficient scientific detail to allow an immediate application. In contrast, it is hoped that the reader will learn from this discussion that certain analytical concepts are appropriate to his decision tasks: to acquaint him with the types or "styles" of models he may choose from, to show him specific examples, and to provide a resource bibliography of modeling literature.

The chapter is organized in four main sections:

—a discussion of the decision types at all levels of educational planning and operation, with an emphasis on those which do now, or will shortly, lend themselves to model analysis

—a set of brief descriptions of actual decision models used by diverse planners and managers in various levels of education

—an assessment of the overall availability, suitability, and potential of education models

—a bibliography of education model literature

Planning and operation of education is a multi-level human service industry, in which many of the decision problems are peculiarly associated with the educational mission and many others are common to the management of all large, complex organizations. Interestingly, a larger body of literature on formal model development is found in connection with the administrative and organizational functions than in connection with teaching methods and instructional decisions. Historically, the "professionalization" of school administration has preceded the "professionalization" of teaching, so that the development of formal administrative models has a head start of many years on similar developments in instruction. The concept of "learning models" is as old as empirical psychology, but the adaption of psychological models of learning to the engineering and design of instruction is of relatively late vintage.

In this chapter, our emphasis is upon models for educational decision making, rather than models for the description of educational phenomena. This emphasis is necessary in order to reduce the number of examined approaches to a practical level; an attempt to include the full range of models used to describe education would involve extensive representations from political science, economics, sociology, urbanology, epidemiology, communication theory, and other disciplines. Instead, the material which follows will address only this question:

> What are the actual decisions and decision problems made frequently at all levels of education, and what models, or model types, are appropriate to these decisions?

Decision Levels

The American educational complex is organized in the form of a pyramid, with approximately 50,000,000 students at the base and the Division of Education of HEW at the apex. As should be expected, the vast majority of decisions, as well as the consumption of resources, goes on in the bottom layers of the pyramid, and, clearly, the time span and scope of decision tends to be shorter at the base than at the apex. As shown in Figure 1 the base is comprised of school districts and other educational agencies (such as colleges and universities), and this base is further divided into campus or sub-district, program or curriculum, and individual student. It is appropriate to show more detail at the base level because, as already indicated, most of the decisionmaking goes on at this level. (Also included in the base is the proliferating number of non-college post-secondary schools and programs, both public and proprietary.)

At the next higher levels in the pyramid are the Intermediate or Sub-state units, and the State Education Agencies (SEA's) or Departments. In many states the SEA or the Intermediate unit (which is typically a county or cluster of counties) actually operates several schools or program—for example, in specialized service areas such as education for the severely handicapped. For the most part, though, Intermediate and State units exist to regulate, lead, and provide supportive services to the units at the base of the pyramid, or to serve as a funding conduit between higher levels of government and the schools and colleges. Typically, planning and decisionmaking at these State or Intermediate levels are with respect to the basic level, and the State or Intermediate plan is "driven" by the goals and programs at the basic level.

Between the State and Federal level are a few multi-state associations or authorities, mainly in Higher Education. These organizations are only quasi-governmental, however, and make relatively few decisions which directly influence practices in schools and colleges.

At the Federal level, most decisions regarding education are made in the Division of Education of the Department of Health, Education and Welfare (although many educational programs exist in other Federal agencies). The Division consists of the Office of the Assistant Secretary, the Office of Education, and the National Institute of Education. The differing responsibilities of these Federal units are still being defined. The authority for operation of Public Schools and regulation of private institutions rests, of course, with the States.

As indicated earlier, the different decisions made by each level frequently differ mainly in scope and timing, with swifter, more detailed decisionmaking as the lowest level. Thus, each level's decisions admit of different forms of mathematical quantification and require different levels of statistical precision . . . even when the parameters of the decisions (e.g. staff requirements) appear to be the same.

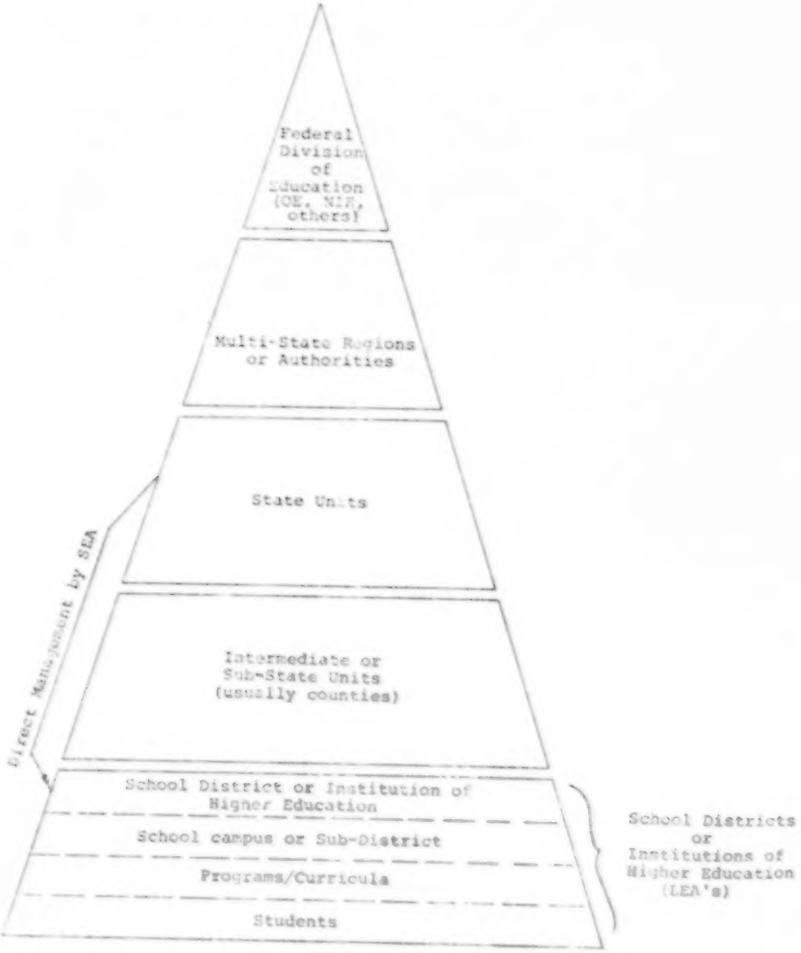

Federal Division of Education (OE, NIE, others)
Multi-State Regions or Authorities
State Units
Intermediate or Sub-State Units (usually counties)
Direct Management by SEA
School District or Institution of Higher Education
School campus or Sub-District
Programs/Curricula
Students
School Districts or Institutions of Higher Education (LEA's)

FIGURE 1—*Levels of Educational Decision*

School District Level: Student-Centered Decisions

Most educational decisions are about what to do to, for, or with individual students. They are typically made by an instructor or classroom teacher, and usually without a formal decision model.

The most familiar educational decision problems involve what to have students do (or what to expose them to) over the next several minutes, hours, or weeks. These decisions include the choice of teaching methods and materials, the making of student task assignments, and the selection of reinforcements to be applied. Ordinarily, this level of decision is based on intuitive models, "in the instructor's head," rather than formal quantified models; this trend is probably attributable to an absence of both theory and empirical evidence supporting ideas of effective instruction.

Averch, et al. [4] have recently released a major study for the President's Commission on School Finance, entitled *How Effective is Schooling?* . . . in which they address critically an abundance of descriptive and evaluative literature on educational innovations and experiments. Searching for generally predictive program characteristics, they are able to find numerous cases in which treatments or program types appear to be effective, but are unable to identify any general or systematic patterns of effectiveness. This finding leads one to conclude that the factors which account for success in both conventional and innovative programs have not yet been identified, nor, indeed, conceptualized in a way that admits of scientific appraisal. In short, there is high uncertainty at this time about what variables should be part of a decision model for optimizing instruction through school management policy decisions.

The areas in which modeling for short-run student decisions do exist are in curriculum engineering [63], and behavior modification (or application of operant conditioning to instructional design [39]). In curriculum engineering models, the course of study is conceived as a branching program of student performance objectives, in which students are routed through consecutive steps on a path which leads to attainment of the major instructional objectives. Such systems are frequently computer assisted (the computer terminal is the instructor) or computer mediated. Behavior modification models are used largely in basic skills training or therapy-like programs to correct behavioral dysfunctions in students. They operate by emphasizing a formal reinforcement schedule, such that positive sanctions and rewards are allocated to student performances according to a specific model, and behavior is thereby "shaped" to conform with educational objectives.

It should be emphasized that, despite the abundant literature in this area, especially in curriculum engineering, the overwhelming majority of short-run student decisions are made without benefit of mathematical models, or, indeed, formal models of any kind.

Longer-run Student Decisions

Longer-run student decisions are usually associated with the guidance and counseling function, which assists the student in course and program selection. Again, these decisions are mainly instinctive and non-formal, although they resemble a goal-programming approach, in which interim decisions are made on the basis of their probability of enhancing the attainment of the highest practical level of student aspiration. Typically, the inputs to this decision model are test scores—intelligence, basic skills, vocational preference, etc. —augmented by preference statements of the student and parents (the student alone, in higher education). The alternative courses of action available to the decisionmakers are constrained by the staff and programs available in the institution, the financial resources of the institution and student, and the belief about the relationship between certain courses and probable goal attainment. Given this configuration, longer-run decisions are highly sensitive to such variables as the completeness of student ability and preference data, the number of alternatives to choose among, and the firmness of knowledge about the relationship between program selection and results. Thus, it is not surprising that guidance and counseling decisionmakers are increasingly employing more ambitious student information systems, as well as matching or searching logic which line up student need with available programs. This approach is most effective in the selection of post-high school programs for secondary students where the plan may be drawn from the approximately 2,500 colleges and universities in the data bank, plus countless other post-secondary programs, and a decision can be made which satisfies student objectives, and constraints of locale, cost, and other variables.

In the lower grades, longer-run student decisions are made less formally, except in the cases of students with diagnosed handicaps or disabilities. In these instances, the plan is based not on empirically verified models of prescription and treatment but on the collective judgment of a team of interdisciplinary specialists (psychologist, audiologist, social worker, etc.).

School District Level: Curriculum Decisions

Curriculum decisions are generally of two types: what curricula (or programs) to offer; how to offer them. The former decision is generally political, in response to demands or preferences from students, parents, and educators, made within the constraints of public law and availability of necessary instructional resources (such as teachers with appropriate skills).

In basic education at the secondary level, curriculum or course offerings are frequently constrained by a minimum enrollment, which will bring program costs into a feasible range [21]. In Higher Education, whole programs and colleges within universities are sometimes evaluated on economic viability, with favor often shown to self-sustaining or even profit-making programs [25].

There are, and can be no, purely rational or deterministic models of what programs should be offered in the educational program—since such a decision is properly in the domain of political values. The closest one might come to such an approach would be the selection of curricula on the basis of aggregating individual student plans (mentioned earlier) so that a program becomes, in effect, the intersection of several student plans.

At certain macro-levels of decision-making, the choice of *curriculum mix* may be influenced by a cost-benefit model, in which the relative economic benefits of alternative programs (e.g., general education versus vocational education) are evaluated by projecting the expected return in student earnings or reduced public expenditures (welfare, for example) and contrasting that return with the expected alternative costs [40]. A less rigid variant is cost-utility evaluation, in which decisionmakers assign common unit weights to the non-economic benefits of planned programs and attempt to maximize the value of the utility/cost fraction [74].

More directly applicable than cost-benefit analysis is cost-effectiveness comparison, in which, after goals have been politically determined, alternative methods or means to accomplish common objectives are compared. Depending on the decision problem the model indicates the least expensive way to achieve a given objective or the most effective method which falls within an upper bound of allowable cost [64]. The ability to use cost-effectiveness decision models is constrained by the limited knowledge of cause-effect relationships in instruction (necessitating reliance on "Delphi" or other soft estimation techniques [74, Appendix A]) and also the obscurities of educational cost accounting, which tends to be superficial and imprecise in relating expenditures to outcomes. (The development of education "production functions" is concerned with solving these analytical problems, and will be discussed again in this chapter.)

School District Level: Campus Operating Decisions

Any school district or university is faced with decisions about the assignment of students, staff and resources on a building or campus basis. These decisions resemble those above, but the emphasis is on specific dimensions of physical space and time.

The problem of *class scheduling* is one of finding an acceptable plan which minimizes unsatisfied student demands, while assuring a maximum utilization of staff, facilities and equipment. A "class," thus, is a set of students assigned to an instructor (or group), in a unique space, at unique times. Because of the dimensions of the problem, class scheduling lends itself to a number of well-developed industrial management models in which conflict matrices are generated, and schedule conflicts are minimized according to quantifiable criteria such as to minimize:

- students not satisfied in course choice
- unevenness in the utilization rate for staff and physical resources
- interruption of sequenced courses

Such models and systems lower conflicts to a prescribed minimum, and there is a growing body of models and software available to perform this function. In addition, several of these systems include features which allow for educational innovation, mainly by allowing for the flexible aggregation of small units of time (usually called modules) into a variety of class configurations. Such techniques allow for a broader range of student choice, and also a wider range of program design alternatives (such as flexible or "modular" scheduling) [24, 28, 30, 33].

Close to half of the basic education students in America go to school on a bus, and most of those on a bus which is operated by the School District. Efficient *vehicle scheduling* also admits of several mathematical decision approaches, based on assignment or queuing models, which can serve to reduce miles traveled, number of vehicles, number of stops, mean travel time for students, etc. When used in a simulation mode, school transportation officials can "try" policies in which, for example, trade-offs are made between the distance students must walk to a bus stop and the net cost of the system.

Current opinion is that the use of well-developed modeling methods in school vehicle scheduling can significantly reduce cost and even minimize accidents [2, 52].

In recent years, of course, vehicle scheduling has been increasingly associated with school desegregation. Vehicle scheduling systems can be adapted to respect constraints on the race-mix in schools (see "Student-assignment," below) or to achieve desired race-mixes, subject to contraints of efficiency and cost. It is possible, for example, to create a linear programming model in which the variables in the objective function are school boundary locations and the objective is to minimize the gap between actual and desired racial distributions, subject to transportation constraints on largest allowable travel time, or mean travel time, etc. [20, 31].

Several modelling techniques may also be applied to the general problems of vehicle maintenance and replacement, the goal being to determine optimum times for the "trading-in" of vehicles, as a consequence of trade-in value, maintenance costs, and costs of replacement vehicles.

Large campuses or school buildings are repositories for large *inventories* of supplies and equipment, some of it consumable, some of it subject to deterioration and dysfunction. (In some cases, a centralized warehouse facility serves several buildings.)

The ongoing assignment of professional and non-professional personnel is frequently aided by information systems which search through records of staff abilities, certifications, and experiences. Mathematical assignment models can be used in conjunction with such an information system to achieve such staff assignment objectives as ethnic mix, sex mix, age/experience mix, and so forth.

Personnel assignment can also be done in conjunction with resource requirement models (to be discussed later). In these cases, the main focus of the model is the estimation of what manpower and materials are required to operate a particular program or project. Several techniques familiar in industrial engineering (such as Critical Path Scheduling) can be used in this process.

School District Level: District-Wide (or University-Wide) Decisions (Short-run)

So far, we have discussed several student-centered, curriculum-centered, and campus-centered decisions. In most cases, these decisions are related to short-run operating problems, with high needs for precision. Clearly, staff assignment in a particular school, at a particular hour, requires a type of decisionmaking different in precision from estimating, say, a university's staff needs for ten years.

At the district-wide, or university-wide level, the shortest planning time frame is usually a year, a period keyed largely to the annual appropriations cycle. For the most part, the short-run decisions discussed below are likely to be made in the annual budgeting cycle. In addition, longer-term decisions are also discussed, and, in these cases, the limit on the number of years to be included in the span of prediction can be as long as the expected life of a facility.

Probably, the most significant educational resource decision is *personnel selection*. Typically, one-year staff planning is based on a forecast of demand (number of students and program/course preferences), an analysis of available staff, and an estimate of new staff requirements. In more sophisticated planning, an expected attrition or turnover rate is also utilized, so that total "positions" by staff type may be adjusted to show total "hires."

"Recruiting" may be a matter of finding staff with scarce abilities (aided by search of a talent bank data base), but, in basic education, is more likely to be a system of selection rules for choosing among the oversupply of available teachers and administrators.

Little formal attention is paid to "pricing" models in staff planning. Generally, decisionmakers *intuit* the relationship between salary determinations and effectiveness of recruiting.

Obviously, any analysis of staff requirements involves student population and enrollment forecasting. In the area of basic education, one finds deterministic models, in which the school population is determined by a set of demographic conditions in the community served by the schools [74].

In Higher Education, vocational and non-public education, however, enrollment forecasting is more complex, since it must include feedback from the course/program offerings. Public elementary/secondary enrollment is determined largely independently of what is offered in the schools; other educational enrollments are affected by pricing, "product line," and competition [73].

Deciding the amount and type of manpower and materials for *maintenance* can be based on several approaches, most of them rule-of-thumb planning factors about the size and age of facilities to be maintained. Most school districts maintain an administrative distinction between maintenance and custodial functions, with the latter providing continuing, loosely monitored service, and the former an *ad hoc* utilization of a pool of staff and other resources. Some school districts are developing more carefully planned programs of preventive maintenance, and using modeling to schedule maintenance activities, but, for the most part, educational maintenance consists of replacing lamps after they burn out and repairing surfaces after they have deteriorated.

District-Wide Decisions (Long-run)

Many educational decisions require more than one or two years to implement, and, for that reason, must be based on forecasts and estimates of at least three years.

In public school districts, where the system is required to provide education for everyone in the community, the *number and location of facilities* is determined largely by long-range population forecasts, for which several mathematical models exist [37]. In higher education and non-public education, the planners have a choice with respect to how many people they will serve, so that demand forecasts are either adjusted downward (if the decision is made to *not* serve the full demand) or upward (if the planners determine to expand student recruiting).

Other mathematical components in the process are:

—calculating deterioration of existing plant.

—computing or estimating the differential costs of starting construction in alternative years (financing rates versus construction costs).

—evaluating extended utilization of existing facilities (by extending the service day or year) [60].

Choice of facilities location is usually made among relatively few available plots of land, but may, in some cases, include the full gamut of site location variables used in industrial site selection models. The variables may include not only the hard data above, but also softer data regarding racial mix, esthetics, convenience, employee preference, and even environmental impact.

In public, basic education, student assignment to new facilities is determined largely by propinquity, and mathematical programming models may be used to minimize distance from school (or transit time), or a simulation model may be devised to accomplish the same purpose.

Mathematical programming or simulation models of student assignment may also be employed in devising districting plans which reduce racial imbalances (see earlier discussion of vehicle scheduling).

District-Wide (Major Decisions)

Decisions about the creation or elimination of major programs are made in a fashion analogous to the selection of curricula (discussed earlier).

Rarely are such decisions made on criteria which are quantified. Whatever cost-utility evaluations are performed are done implicitly through political negotiation.

All educational organizations have a need to *estimate revenues* in the short- and long-run, although many are reluctant to project multi-year income. Because most educational organizations generate at least some of their own revenues (by levying taxes or tuitions) it is necessary to project the revenue from all non-controlled revenue sources, so that the appropriate rates and schedules for one's own revenue sources can be determined.

Public education revenue analysis, given the current financial system, generally involves prediction of property tax bases and the inflation of their assessments. State subsidy revenues involve, also, enrollment forecasts, staff forecasts, and other variables, depending on the state. In recent years, school districts must also estimate non-formula income from state or Federal categorical funding programs, whose year-to-year funding levels are often quite difficult to project [74].

Still another area of revenue source decisions involves the design of taxing plans to raise educational revenues. If models are used, they entail complex relationships in which the base of the proposed tax must be forecast, and alternative rates must be simulated to find a rate which will produce the needed money without having a negative impact on the base (e.g. liquor, cigarette taxes). In recent years, non-tax revenues must also be designed and analyzed; in general the prediction of income from lotteries and other gambling is even more difficult than from conventional revenue sources.

Intermediate (County) Level Decisions

The education intermediate unit—usually a county or cluster of counties—is concerned with the public school districts in its region, the public two-year colleges, and some vocational technical programs. While such units frequently operate their own educational programs—often in the area of services for handicapped students—they are more likely to be regulatory or supportive of the smaller educational units in the territory. Thus, their decisions are of two main kinds: program and support.

To the degree an intermediate unit operates *its own educational programs*, it must make decisions parallel to those made in the university or school district: student plans, curriculum, operations, scheduling, maintenance, etc.

Intermediate units must frequently be concerned with transportation scheduling, construction, and the full range of activities found in the more basic units, and, thus, have parallel opportunities for the use of decision models.

A major planning problem is the creation and allocation of *supportive services to districts* in the region. In the current era of IU's, these services tend to be: data processing, instructional materials loan, workshops and other staff development, bibliographic and abstracting services, proposal development assistance, liaison with state government, and general education consulting. Because the creation of such units tends to stimulate demand for them, the intermediate unit managers are frequently faced with complex short-run allocation problems for their own small staffs and computer facilities. In these settings, quantifiable criteria can be used to ensure a desired allocation of services, and the criteria usually are related to district size, wealth, grade-span, urbanicity, or other statistical parameters.

State-Level Decisions

The state-level education unit is quite difficult to describe, because it exists in so many varieties and is usually in transition.

In some cases, the State actually operates basic institutions (such as state colleges or state schools for the mentally retarded); in these cases its decisions parallel those made in school districts or universities. For the most part, however, the state education agency regulates schools, school districts, and institutions which it does not directly operate. Thus, an explication of state level decisions requires description of the relationships which may be obtained between the state and the operating units.

Notice that the three general roles for the state agency are *regulation, leadership,* and *service.* That is, the state may enforce laws and codes, it may establish statewide goals and priorities, or it may provide supportive services (like the Intermediate Units) to supplement the resources of the operating units. The most significant area of service, of course, is the providing of substantial funds for the operating units (usually 20 to 50% in public school districts). But states also *regulate* the local budgeting and accounting of the operating units, and may even exert leadership in recommending new resource allocation policies or expenditure strategies.

The SEA tends to exercise its three roles in regard to six functional areas: finance, students, staff, curriculum, facilities, school-community relations.

It is easy to conceive decision problems in each of the eighteen role/function areas. For example:

Finance

Regulation—How must school districts handle their accounting and auditing?

Leadership—To what purposes should local and state resources be applied?

Service —How much money should be provided to LEA's and colleges?

Students

Regulation—What students must be served?

Leadership—What students ought to be served?

Service —What resources should be provided to augment services to special needs students?

Staff

Regulation—What certification requirements must be met?

Leadership—What staff types should districts be encouraged to find?

Service —What training should be provided to upgrade the abilities of existing staff?

It is clear how parallel questions may be raised in each area. In the sections below, however, the most important decision problems—and their amenability to decision models—are discussed in more detail.

The most controversial and complex area of state decision making at present is in *determining appropriate subsidy formulas for organizations*, institutions, and, in some cases, individuals in the state. While there has been much recent attention to the construct of "equality" in support of education, the notion of equitable distribution of state/local burden has been a main consideration throughout this century [14].

Determining state subsidy approaches requires a trade-off between *equality* and *equity*. A straightforward levelling of state support will penalize those communities with limited local resources and produce certain inequities. A preferential treatment of less wealthy districts, however, is often perceived as inequitable by wealthier communities, since many of them also feel their local tax burden to be uncomfortably large. (An interesting statistical peculiarity is that *any* across-the-board increase in state subsidy will reduce the correlation between local wealth and school expenditures.) State subsidies to educational agencies are generally based on a mathematical formula involving enrollment (adjusted to reflect average daily attendance, or attendance of "weighted" pupils), and a per (weighted) student subsidy rate. In some cases, the subsidy is directly tied to the number of students; in other states, the subsidy is tied to an allowable number of teaching positions (as a function of average attendance), with the state providing part of the teacher payroll. In those formulae concerned with equalization, certain communities receive a premium if their taxable wealth (assessed property valuation/student) falls below a state standard; this subsidy is a function of the gap between actual taxable wealth and the criterion, as well as the *local* tax rate—to insure that local communities will exert reasonable "tax effort."

Certification of educational employees is both a quality and quantity control device. That is, it can be used both to ensure that education personnel meet certain minimum skill and preparation standards, and also to arbitrarily limit the size of the manpower pool.

Many states are currently engaged in improving the definition of certification requirements for educators, in order to make them more "performance-oriented," rather than oriented to "credit hours" of training. Clearly, increasing certification requirements and introducing more specialized certifications has the effect of increasing students and revenues at teacher-preparation institutions—most of which are state-operated. They also tend to raise the salaries of educational personnel, because salaries are tied to degrees and "credit hours." It is not surprising that a major political influence on state certification requirements has been the associations and unions which represent education professionals, as well as the administrators of education programs in teacher-preparation institutions. The decision making problem remains, however:

> What certification regulations shall a state adopt and enforce?

In addition to certifying staff, most states periodically *evaluate or "approve" schools and school districts*—frequently as a basis for continued or modified financial support. There are at least three conceptually distinct approaches to the problem:

—*Performance Measurement:* Perhaps the least frequently used approach is to evaluate the student product of the school or district, by assessing the skills and post-school performances of the students. While many states have state-wide testing data at the basic and higher level, few use this information as a basis for evaluating the schools.

—*Operating Characteristics Measurement:* The most typical approach is to use a checklist of salient agency characteristics, such as staff/student ratio, library holdings/student, ratio of certified to non-certified personnel, space and specialized equipment, breadth of course offerings, etc. The assumption behind the use of this statistical and anecdotal data (used in most high school approvals by regional accreditation organizations) is that these characteristics are predictive of educational quality and efficiency. There is, of course, no sound empirical evidence supporting such input-output relationships, and, for that reason, the checklist approach has begun to lose favor in some states.

—*Meta-evaluation:* An approach currently receiving some attention is the meta-evaluation, in which the purpose of the state's investigation is to see *that the local unit is evaluating itself competently*, and to provide guidance and assistance in the evaluation process. This approach necessitates agreement on the allowable range of evaluation methods (and standards for employing them), but this agreement is more politically and technically easy to reach than a set of output standards would be.

There is much activity in standard-setting and evaluation in SEA's. Little of it, however, employs decision models.

SEA's are also frequently required to *define the boundaries of new school districts, or to design consolidated districts*, or specify the allowable location of new schools or college campuses. Among the parameters in such decisions are:

—projected populations and population density.

—degree of contiguity with existing governmental boundaries

—economies of scale versus dysfunctions of excessive size

—feasibility of revenue base

—patterns of racial and economic segregation

A related decision problem, mentioned earlier, is the planning for location of intermediate or regional units of the SEA. In this case, a further important parameter is closeness of match with other regionalization plans of other human service agencies in state government.

Federal Decisions

Like the State Education Agencies, the Division of Education, HEW, has leadership, service, and regulation functions—the last confined to those programs it funds and also to enforcement of Federal laws and judicial decisions affecting schools and colleges. In the late 1960's the Federal government was directly administering the Elementary and Secondary Education Act, as well as other programs, and dealing directly with thousands of school districts and colleges. While the higher education contacts have persisted, in recent years the Federal level has transferred most funds and decision prerogatives to the SEA, and, in fact, does most of its communicating with local educational agencies through the SEA's.

The policy making at the Federal level tends to be of the broadest scope and concerned with the largest planning horizon of all levels in the system. And, if there is a consolidation of several Federal programs into Education Special Revenue Sharing, it is to be expected that the Federal role will be more and more detached from short-run local decisions.

The Federal government's main decision problem is to define its role in the American education complex, most significantly its financial role. Even with the advent of revenue sharing, it will continue to fund programs and projects it judges to be in the national interest, but a larger part of its funds will be allocated at lower levels in the system. There is no basis for determining the Federal role, save the Constitution, which provides constraints on allowable roles.

An important series of subsidy and grant decisions may face the Federal government, if, in the next few years, it is determined that Federal support is necessary to achieve equitable funding and financing of the public schools. In this connection, and in all other decision areas affecting the Division of

Education, mathematical models will be helpful—even if imprecise. Among the analytic techniques which have been and will continue to be, useful are:

- —enrollment forecasts for higher and basic education [37, 47]
- —manpower needs projections [6, 55]
- —models for estimating future Federal, state, and local revenues
- —models for projecting the size of special populations, including minority groups
- —econometric models relating investment in education and manpower development to economic growth [23, 55]
- —models relating educational attainment to expected economic and social benefits for typical individuals [13, 26]

A key educational concern of the Federal Government in general is the relationship between education and the entire economy. Not only do specific federal programs have economic impact (like the National Student Loan program), but also the entire education complex has an impact on the growth of the economy and federal tax base. While ours is not a nationally controlled education system, the Federal government may utilize macro-models developed for nations which control all education. The Organization for Economic Cooperation and Development (OECD) has developed several mathematical models which, for example, relate educational plans to industrial and social development, and which simulate the flow of groups of persons through the education system and the labor market. Such approaches are not directly applicable in this country, but have been used in adapted forms in state-level education decisions. (See later discussion on models for optimizing education and the economy.)

In addition to the above there is an emerging need for models which will assist in managing R&D allocations. The newly created National Institute of Education will confront complex decision problems analogous to those faced by other major Federal Institutes.

II. ILLUSTRATIONS OF DECISION MODELS FOR EDUCATION

Introduction

In the sections that follow, twelve different decision models are described briefly. The examples chosen are not exhaustive but are a good representation of the range of analytical approaches currently being used, and the typical decision settings in which they often are applied. The illustrations include several levels, from nation-wide forecasting, to specific problems in the school or college, to detailed student assignment. They incorporate deterministic and probabilistic models, heuristic approaches, simulators, and optimizers. Each is presented with a minimum of obscure technical information; the goal is to show what they are, how they work, and for whom they might be useful.

The illustrative models are presented in descending level, from nationwide decisions to the local classroom. The levels and models are:

National Level:

- Macro-economic planning
- Forecasting college enrollments and financial aid requirements

State Level:

- Allocation of vocational funds
- Input-output evaluation of local schools
- Forecasting state-wide higher education demands

School District Level:

- Resource Requirements Prediction
- Vehicle Scheduling

A Theoretical Model for Optimizing the Economy and Education

The first model to be discussed will be treated more generally than the remainder in this chapter, partly because it is a theoretical decision model, rather than an actual, useable management tool, and partly because it contains elements which will be detailed in subsequent discussions. Thus, the concepts in Benard's general optimization model for the economy and education [10] are presented without its elaborate mathematical notation.

The Benard model addresses a question of optimum investment policy at the national level. Interestingly, it was developed in France, where the school system is nationally managed, and national economic planning includes direct decisions about the nation's schools. (According to Benard, in France 80% of the schools are public, and the remainder are subsidized by the government.) Beginning with the most general question—what proportions of national "physical" resources should be invested in each sector of the economy—a more limited (though still quite global) question is formulated:

> If Education is regarded as a sector of the commercial economy, how shall investments be distributed between education and non-education activities, so as to optimize economic growth, subject to certain constraints?

In this conception, the economy has a main output: "goods and services," expressed in a dollar value. Education is viewed as a part of the economy, namely that part which produces skilled workers for business and industry. (The other educational outputs and benefits are mentioned by Benard, but play a minor role in the model.) The output of the schools, thus, is expressed in numbers of "skilled" graduates, who, in turn, flow into the other sectors of the economy, or back into education as teachers or administrators. Clearly, different investment policies will produce different numbers of skilled graduates.

An essential concept in this model is the notion of human flow, both through the various levels or cycles of the educational program and through the levels of the commercial sector (that is, levels of skill). An important characteristic of the model is the logic for flows from level to level over time.

During any time period t, at any level (education or otherwise), people will have flowed from one of three conditions in the just preceding time period:

—at the just lower level

—at the same level

—from outside the system (either outside the country, or, if education is viewed as a separate system, from education into the commercial sector).

Similarly, at the end of time period t, they will either move up a level, stay at the same level, or leave the system.

It is possible, therefore, to develop a representation of the human flow through all levels of the education and commercial sector, with calculated transitional probabilities from level-to-level, so that a change in any one level will produce calculable systematic changes in all other levels.

Skilled manpower is viewed as an output of the education sector, in this case, a linear function of the quantity of physical goods and services invested in education. (Factors causing enrollment are considered, in this model, as exogenous variables: investment is presumed to have no influence on enrollment itself, only on the quality [in skills] of the graduates.)

In contrast, investment in goods and services in the commercial sector also produces output in goods and services, according to an empirically determined input-output matrix, which is projected in terms of the impact of technological change on current productivity.

The choice, then, for policy-makers, is between indirect investment through education and direct investment in industry, knowing that under-investment in education will eventually cause counter-productive manpower shortages in industry, along with reduced earnings potential for citizens.

The model is a linear programming model. The output is maximized for each of a series of time periods, $t_1 - t_n$. (Of course, the statistical values of parameters are recomputed periodically.

and the model is re-run before the end of the "global" time period.) Each optimization is also a function, of course, of constraints:

—two constraint vectors on maximum consumption, so that the total output in goods and services does not exceed that which can be consumed by the projected population.

—two constraint vectors on production capacity, so that the input of skilled manpower and investment does not cause industry to grow or produce faster than is possible.

—Three constraint vectors limiting the transition of persons from level to level, so that, for example, the number of skilled persons flowing into level 1 does not exceed the number of openings at that level.

—A minimum growth rate for education, that is, a kind of "safety clause" which ensures that even if the economic calculations justify no, or very small incremental investments in education, there will nevertheless be some guaranteed investments in its cultural and nonquantifiable benefits.

—A budgetary ceiling to reflect political constraints on the maximum allowable investment in education.

The model is intended to allow policy makers to "maximize the present discounted value—over all periods considered—of the sum total of the variables representing the standard of living of the population, of the production potential of goods and services, of skilled labour and of education continuing beyond the time limit (horizon) of the model, and so enable the process of growth to continue beyond that limit [10, p. 226]." Obviously, the various components of the output must be weighted according to some preference function or utility measure, and this may prove to be the most sensitive or politically difficult part of using the approach.

The Benard approach has, in fact, been adapted in an actual model by the Organization for Economic Cooperation and Development in Paris [54]. It has also been adapted for use in statewide educational planning (the Education Coordinating Council of the State of Oregon) and it appears to have influenced many other model-makers, including McNamara [48] (whose vocational education allocation model is discussed later in this chapter). The model has considerable conceptual richness, and invites applications even though calculating the statistical values of the various flow and productivity matrices may prove quite difficult.

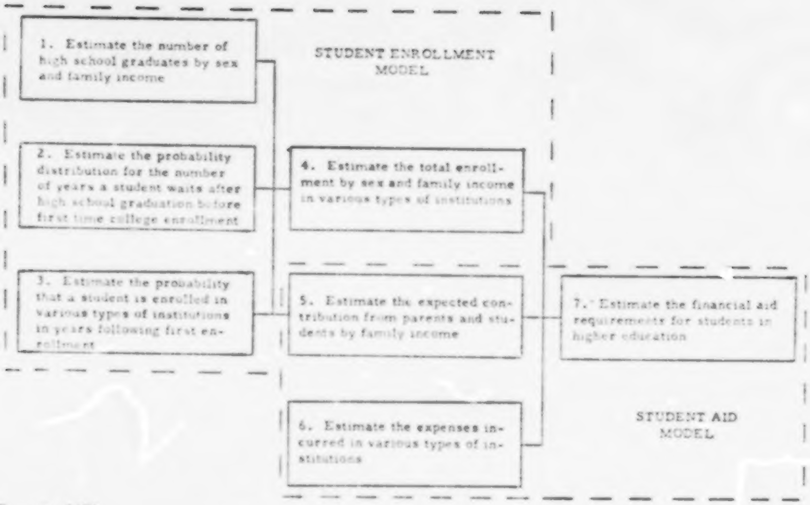
Forecasting Nationwide Undergraduate Enrollments and Federal Financial Aid Requirements

In 1971 the U.S. Office of Education developed a system of probabilistic and deterministic simulation models, whose interim output is a probable projection of the total undergraduate enrollment in the United States, and whose ultimate output is a determination of the amount of Federal aid necessary to support that population, and the distribution of aid types (loans, grants, work-stipends, etc.) [47]. Enrollment projections are expressed as number of students, by sex, by institution type (2-yr., 4-yr., university), by private or public, for each year in the forecast. These projections are subsequently input into the aid calculation model, as summarized in Figure 2. (The same project also produced procedures for forecasting post-baccalaureate enrollments and aid, but these are not discussed here.)

As indicated in Figure 3 a potential student can be in one of three conditions: a high school graduate, a first year student in college, a student in a stage beyond his first year. Students, moreover, are categorized by sex and family income group, data which is both descriptively interesting and an input to the aid calculation.

The sequence of calculations in the series of models is as follows:

1. Number of high school graduates, by sex and income group, for each year in the forecast is projected.


1. Estimate the number of high school graduates by sex and family income
STUDENT ENROLLMENT MODEL
2. Estimate the probability distribution for the number of years a student waits after high school graduation before first time college enrollment
4. Estimate the total enrollment by sex and family income in various types of institutions
3. Estimate the probability that a student is enrolled in various types of institutions in years following first enrollment
5. Estimate the expected contribution from parents and students by family income
7. Estimate the financial aid requirements for students in higher education
6. Estimate the expenses incurred in various types of institutions
STUDENT AID MODEL

Source: [47]
FIGURE 2—*Summary Flow of the Model*

2. An estimate is made of the probable number of years of delay between graduation and starting college for each sex/income group.
3. First time freshman enrollment is calculated from the first two steps.
4. Next the retention or attrition probabilities are estimated. The probability of a freshman continuing to consecutive years is different in each year, and the existing data shows different rates for the various sex/income groups.
5. A distribution of students by type of institution and "control" (private vs. public) is introduced.
6. The enrollment forecast output is produced by consolidating the steps above and reporting the enrollment by sex, income group, institution type and control type in each year of the forecast.
7. To use the enrollment projection as input to the aid model, the next step is estimating the parental and student contribution to college expenses, which is a function of income, number of dependent children, number of dependent children attending college, student assets, and summer income.
8. College costs are estimated by type and control of institution, including tuition, fees, room, board, books, and miscellaneous. Cost estimates are differentiated for resident and commuting students.
9. The final calculation measures the gap between the expected capability to contribute and the projected costs.

The difficulties entailed in using this model stem largely from the number and complexity of sub-models required to perform the various forecasts and estimates. Referring to Figure 3, it should be noted that there are seven main components in the model, and all but the seventh involve at least one of 33 forecasting sub-models.

It is clear that the projections from this model, because of the compounding of errors in the several sub-models, should not be treated as overly precise. Recall, however, that its purpose is to project, in broad terms, the long-range financial consequences of college participation. We can assume that its accuracy in forecasting these variables is significantly greater than any intuitive, or less complete, prediction procedure currently available.

A Model for Allocation of Vocational Education Funds

A recurring decision problem for State Education Agencies is the allocation of a large pool of program funds across a set of programs and projects in local education agencies (LEA's). McNamara has developed a linear programming (LP) model for use by the Pennsylvania Department of Education in allocating Vocational Education monies [48]. (McNamara has also developed several review essays on the application of mathematical programming to educational planning: see [50]).

In McNamara's model, the objective function is:

$$\text{Maximize } z = \sum_{i=1}^{m} \sum_{j=1}^{n} X_{ij}$$

where

X_{ij} = the output in total students of vocational program i from county j.

i = 1, 2, . . . m (vocational-technical education programs)

j = 1, 2, . . . n (counties in Labor Market Area)

A first constraint in the model is *capacity*, the rate at which new students can be added to existing facilities, or the rate at which new resources can be deployed. For this reason, each county/LMA output must fall between the minimum "desired" increase in capacity of program i in county j, t_{ij}, and the maximum "desired" increase in program i in county j, $\bar{t}_{ij}$:

$$t_{ij} \leq X_{ij} \leq \bar{t}_{ij}$$

Notice that the use of "desired" constraints allows for county level judgments, as well as hard data on the number of student stations available.

Importantly, this model is also constrained by budget limitations. Thus, any optimal solution must satisfy the constraint that

$$\sum_{j=1}^{n} h_i r_{ij} X_{ij} \leq H_i \quad i = 1, 2 \ . \ . \ . \ m$$

where

h_i = the fixed incremental cost for each student added to program i in n counties (Note: Common programs cut across counties; this incremental cost per student is a program-specific input).

r_{ij} = the ratio of the students in program i in county j necessary to produce the optimal number of graduates in the program (Note: this ratio is greater than or equal to 1.0, and is the reciprocal of the expected drop-out or failure rate for the program).

H_i = the total marginal increase in Vocational Education funds from the State. The model allows simulation by using several different values of H_i.

The optimum State allocation is a consequence of assembling sub-optimum plans in each Labor Market Area (LMA), each of which is a unique set of counties. This X_{ij} is a marginal increase in the number of students the state wishes to support in each county of each LMA. There is an important constraint, however, on the number of students in a county/LMA, namely, the upper limit on the number of graduates, from each occupation specific program, which can be absorbed by the Labor Market area manpower projection. Maximizing output also requires minimizing the percentage of unmet labor market needs (based on a projection of graduates by year) and also minimizing the occasions on which school output exceeds the manpower opportunities.

This model, when accompanied by appropriate sub-models which generate the various forecasts necessary to the LP model, generates a recommended plan, in which each program in each school district is expanded by an optimum number of students (that is, a number which minimizes unsatisfied manpower needs). The plan, and budget allocation scheme, is for the *next* year, but also has a submodel which projects the ith year on the basis of the first year plan. While typical SEA allocations are made annually, this multi-year extension capability provides further data about decision consequences to better inform the short-run decision.

According to the author, there are some limitations on the use of the model. First, the optimal solution is a function of the supply-demand relationship, in which demand is *manpower demand in a given Labor Market Area.* The model assumes that first jobs for graduates will be in the same geographical area as the school—an assumption that the author argues is relatively reasonable. Second, the manpower demand is not offset by the "social demand," that is, what programs *students want to take.* It appears, however, that the construct of "minimum desired increase in student enrollments in a program" provides a mechanism for inputting social demand, even if that is not exactly how its author construes it. Third, and related, is that the resource allocation model removes monies from programs which oversupply the LMA demand. McNamara indicates, however, that this constraint may be relaxed by the policy maker and does not, therefore, impede a more "social demand"-oriented planner from achieving his own planning objectives.

Perhaps the most difficult data problem in the model is the projection of LMA outputs from non-public schools. As currently designed, the model assumes that the state supported programs will satisfy the entire unmet demand. This problem can be solved through closer scrutiny of the output and projections of private and proprietary vocational programs.

This model is clearly one of the more advanced approaches to resource allocation in education, made possible in part by the more easily quantified outputs of vocational education. Its usefulness in both longer-range planning and short-range allocation decisions is considerable.

A State-Level Input-Output Planning Model for Local Educational Agencies

The New York State Education Department has developed an ambitious project known as Performance Indicators in Education (PIE). As part of the project, SEA staff have attempted to develop the most elusive of education models—the production function: a mathematical model which relates input and environmental variables to performance outputs of local school districts [68]. The developers have taken an especially interesting approach: instead of attending directly to the effectiveness of specific educational programs and treatments, they emphasized the large set of non-instructional variables (including student and community characteristics) which, according to most research, have a greater predictive power than typical instructional variables (such as method or class size). Given this view, the model is used to calculate an *expected* performance for a school district, given its characteristics. To the degree that a given program or treatment exceeds or falls short of the expected performance level, the program is judged to be effective or ineffective. The model is built on a simple conception of the educational process. The concept argues that the output of the schools is a function of student characteristics and environmental conditions, as well as school processes.

The most problematical variable is school processes, and effectiveness is defined as the difference between what is expected given all the non-school inputs, and what is observed. A particular version of the model considered the following as inputs, conditions and outputs:

(a) *Input*

Mean of 1967 and 1968 district means in first grade readiness tests (1R67,68)

Mean of 1967 and 1968 district standard deviations in first grade readiness tests. (1Rsd67,68)

(b) *Surrounding Conditions*

FTV = Full Tax Valuation of district, divided by enrollment

PR70 = Proportion of Minority Group Students in third grade in 1970

S68 = "Size," enrollment in grades 1–12, divided by 1000

D68 = "Density," square miles in district divided by number of pupils

(c) *Output*

3R70 = Third grade reading mean for 1970

These variables—which are only part of the list developed—are taken as indicators of the input, environment and output. By examining historical data on the schools in New York State, the model makers have been able to perform regression analyses, in which the relationships of the variables are quantified. Equations are in the general form

$$OV = a\,(V_1) + b\,(V_2) + c\,(V_3)\ .\ .\ .\ + t\,(V_n) + k$$

where

OV = value of the output measure

$V_1, \ldots, V_n$ = input and environmental values

$a, b, c, \ldots, t$ = regression coefficients

k = a constant

In the case cited, the general equation becomes:

$$3R70 = .298\,(1R67{,}68) - .112\,(1Rsd\,67{,}68) + .030\,(FTV68) - 5.221\,(PR70) + .13\,(S68) - 1.614\,(D68) + 14.75$$

(This equation accounts for .332 of the variance in actual scores.)

Thus, for any specific set of values for a given school district, one may calculate an expected value (and variance or range of probable values) for "3R70," the third grade reading scores of the district.

The PIE model is an advance on the most elusive of educational problems: the relationship between input and output. There are objections, of course, to any set of output measures, and, indeed, the developers admit that the variables used may not be the most sensitive or best predictors. Importantly, users of this or other systems must never treat the analyses as more precise than they are, never confusing a fixed value with a range of values, never imputing significance to non-significant differences. In the authors' words:

> "The PIE System is designed to reduce the element of chance in decision making. . . . Its immediate value will be to help local districts recognize those program areas in need of more detailed evaluation and perhaps additional resources. The system will eventually allow the development of simulation models for predicting the consequences of decisions before they are made."

Clearly, the developers of PIE make no claims for the system beyond those which are supportable in the data.

A Model for College Enrollment Forecasts

There are two relatively distinct problems in forecasting enrollments in any basic or higher education program:

a. Projecting the number of persons entering the programs or system, and

b. Projecting the distribution of students at each level (usually years) in the program or system.

The first problem usually entails modelling the relationships between the principal education system and some other educational system or non-educational system. For example, in basic education, it is necessary to relate the number of incoming kindergarten or first grade students to the demographics of the community (e.g., birth rate, housing changes, etc.); in higher education, the main predictor of incoming freshman is a mathematical model of the impact of high school completion on college enrollment. In the case of projecting kindergarten students, it is possible to build a deterministic model, because there are direct legal connections between the census for certain age cohorts and the obligations of the schools.

In the other cases, however—grade-to-grade transition in basic education, basic-to-higher education, or year-to-year transitions in higher education—the process is stochastic rather than deterministic. The process is described simply in Figure 3.

Only some of the branching possibilities are shown, but they illustrate the principal concept. Notice that one of the possibilities for high school graduates is going to college; the transitional probability for going to college is TP_{ic}. In the next year, one of the possibilities for college freshmen is to become college sophomores; this transitional probability is $TP_{i+1,\ c+1}$. Clearly, if the transitional probabilities are known, an estimate of the enrollment at any one level, in any year, allows a forecast of enrollments at higher levels for subsequent years.

Oliver [53] has discussed the adaptation of this approach for predicting gross enrollments at the University of California at the state level. In one model, he describes, enrollment is based on the fraction of students moving from level to level, and number of students coming in at something other than the beginning level. The reader is referred to [53] for the mathematical development.

Enrollment forecasting is possibly the most important input to long range planning and resource requirements analysis. Illustrations of its relationships with other modelling problems appear in the next sections of this chapter.

A Model for University Resource Requirements Analysis

Hussain [25] reports on an ambitious resource requirements prediction model (RRPM) for a university, developed by the National Center for Higher Education Management (NCHEM). This ex-

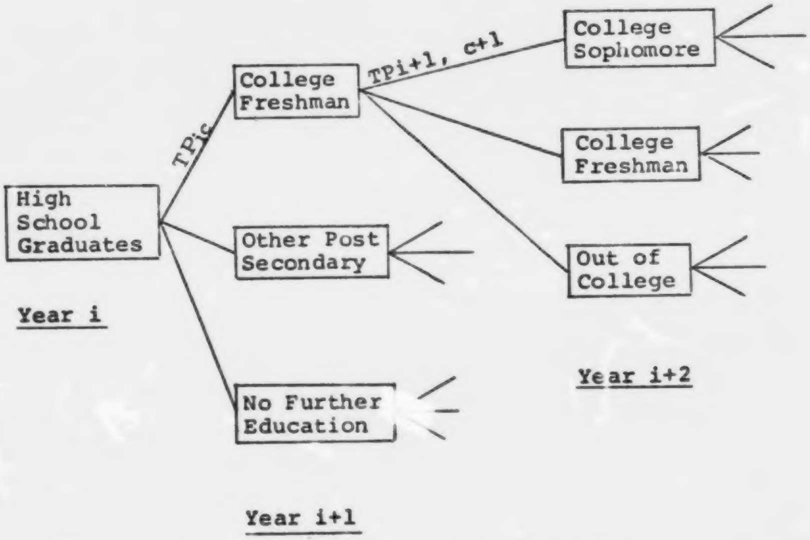


FIGURE 3—*Simple Transitional Model for Three Years*

tensive projection system goes through four analytical phases:

Instructional—calculation of all direct instructional resources (staff, space, etc.) and expenses for faculty, supplies, travel, and equipment, by discipline or department.

Unit Cost—calculation of unit instructional costs, both direct and indirect (allocated), by field of study and student level.

Non-instructional Program—calculation of costs by sub-program, for all sub-programs which are not instructional or academic support.

Space—calculation of space requirements and construction costs.

The principal input to RRPM is an estimation of the number of students to be enrolled in each major field of study, at each undergraduate level, for each year of the plan. From this data, and an analysis of the historical distribution of courses and resources for each field of study, in each level of schooling, NCHEM has generated the Induced Course Load Matrix (ICLM).

The ICLM is essential in the projection of student credit hours (SCH), weekly student hours (WSH), faculty load (FCH) and full time equivalent staff members (FTE). A portion of the model is concerned with projecting direct instructional expenses. This submodel corresponds to the Instruction phase of the process.

The fourth phase, Space, addresses a critical problem for most expanding universities, namely the projection of facilities requirements and attendant construction costs. This process involves taking data from the earlier resource requirements projections and distilling the incremental increase in space requirements (in each of 22 defined space types). As with other resources projected, the model calculates the quantity of resource requirement (space, or staff, or equipment) and multiplies by input price schedules and estimates.

RRPM is a *deterministic* simulation model, used to predict the consequences of *given policies* (such as faculty load), under conditions of different external demand (enrollment by field of study). Thus, like other resource requirements models, its main function is to aid decision makers in testing the consequences of current or contemplated policies. Because its accuracy is constrained by external variables (such as the student forecast or the ICLM distributions) its utility will vary from setting to setting.

As currently designed, RRPM can be used in universities with as many as 90 Fields of Study ("Majors"), and can accommodate 7 levels of students, 5 types of faculty, 4 non-instructional staff types, 4 different modes of instruction (each with its own space requirements), and, as already mentioned, 22 space types. As such, it is a highly flexible planning resource, particularly for those institutions which have adopted the NCHEM program-structure.

A Model for University Facilities Analysis

Sisson, Jacobson, and Robinson [60] have developed a Facilities Analysis Model (FAM) for use by the California Coordinating Council on Higher Education (CCHE). The main objective of using the model is to develop mathematically defensible standards for facilities utilization (mainly hours of service and scheduling of facilities), so that the college may operate "near the least cost point." As is apparent in Figure 4 there is a Total Costs Curve, which is a complex function of the Operating Costs Curve and the Capital (construction/equipment) Costs Curve. Intuitively, it is clear that heavier utilization of existing facilities increases operating costs, while it decreases capital costs. The mathematically interesting questions are:

—How do these two independent curves behave?

—What institutional characteristics are sensitive variables in influencing the behavior of the curves and, thus, the value of the lowest Total Cost point?

The policy-related outputs of the model are, first, a set of utilization

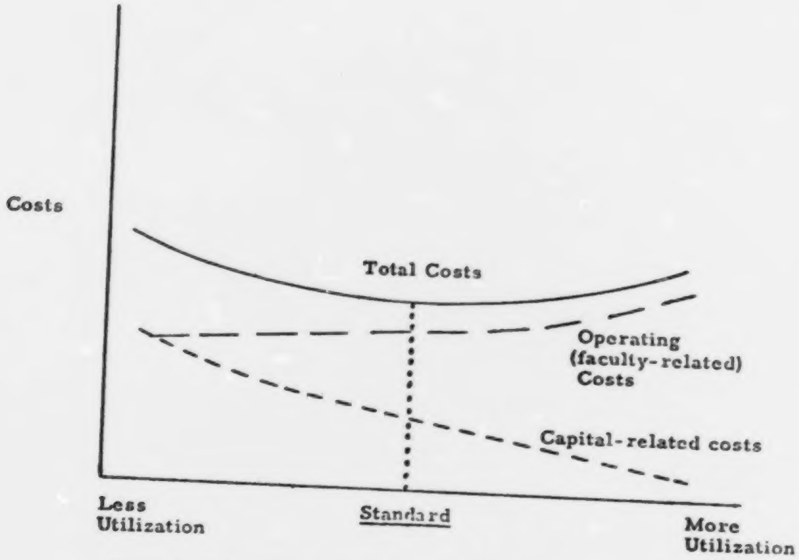


FIGURE 4—*Relationship Between Operating, Capital, and Total Costs*

standards (justified by least cost analysis) and an identification of policy-related institutional characteristics which offset current implementation of the standards, or influence the determination of standards in subsequently developed colleges and universities. Among these factors are:

—The age of the institution (and its rate of growth)

—The gap between enrollment and maximum capacity

—The client groups served, classified by (1) day students, evening students, (2) full-time, part-time, (3) discipline or program, (4) degree, certificate, transfer or non-degree

—Location of institution: urban, suburban, rural.

Recall, that the FAM is *not* intended as a continuing scheduler for facilities, rather as a policy formulating resource, to be used only as often as the general policies on facilities utilization are to be evaluated and revised by the user. Each time it is used (which may be only once in a given college or higher education system) it is run several times to calculate the impact on total costs of alternative policies. The analytical flow of *one* iteration of the model is shown in Figure 5. The computations and analyses are:

1. Forecast enrollment by program or "major" and, using historical data, calculate the demand for classes (weekly student hours) in each major.
2. Schedule the total number of class hours over the course day (the FAM treats all days the same), assuming policy i about hours of utilization. Deduce from this approximation of the master schedule the demands on facilities. Adjusting for any input policy on percent of utilization, calculate the space needs required to fulfill enrollment expectations (the basis for capital costs).
3. Deduce the faculty contact-hours from the scheduled class hours; using the distribution of courses (by discipline) within each major, deduce the number of FTE faculty members required by department, building in assumptions about faculty work-loads (the basis for estimating faculty operating costs).

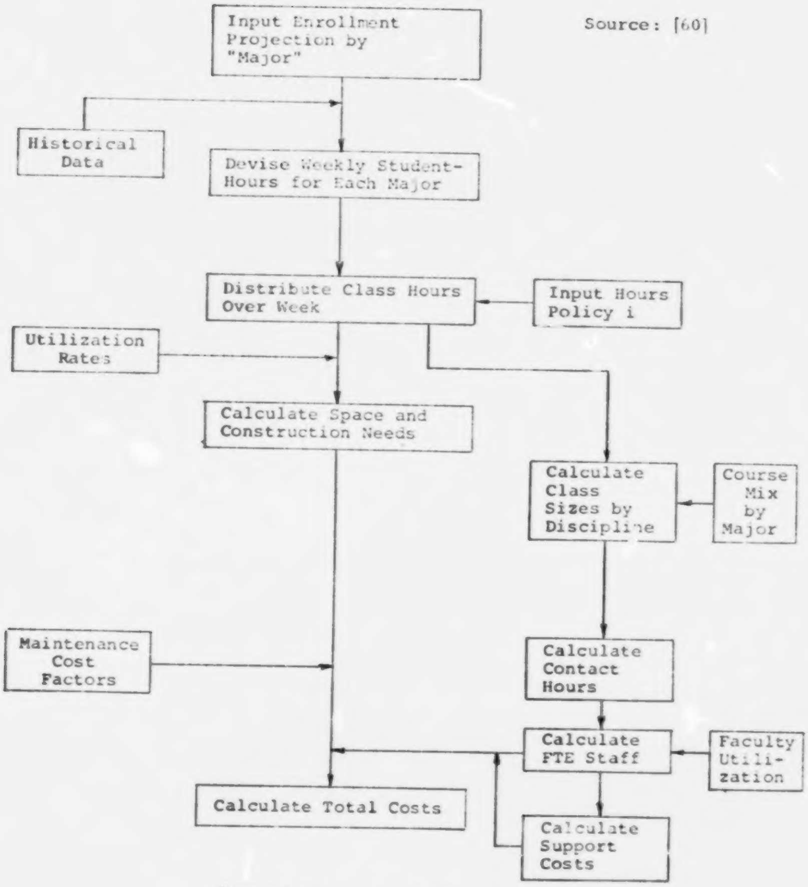

Input Enrollment Projection by "Major"
Source: [60]
Historical Data
Devise Weekly Student-Hours for Each Major
Distribute Class Hours Over Week
Input Hours Policy i
Utilization Rates
Calculate Space and Construction Needs
Calculate Class Sizes by Discipline
Course Mix by Major
Maintenance Cost Factors
Calculate Contact Hours
Calculate FTE Staff
Faculty Utili-zation
Calculate Total Costs
Calculate Support Costs

FIGURE 5—*Flow of the Facilities Analysis Model*

Source: [60]

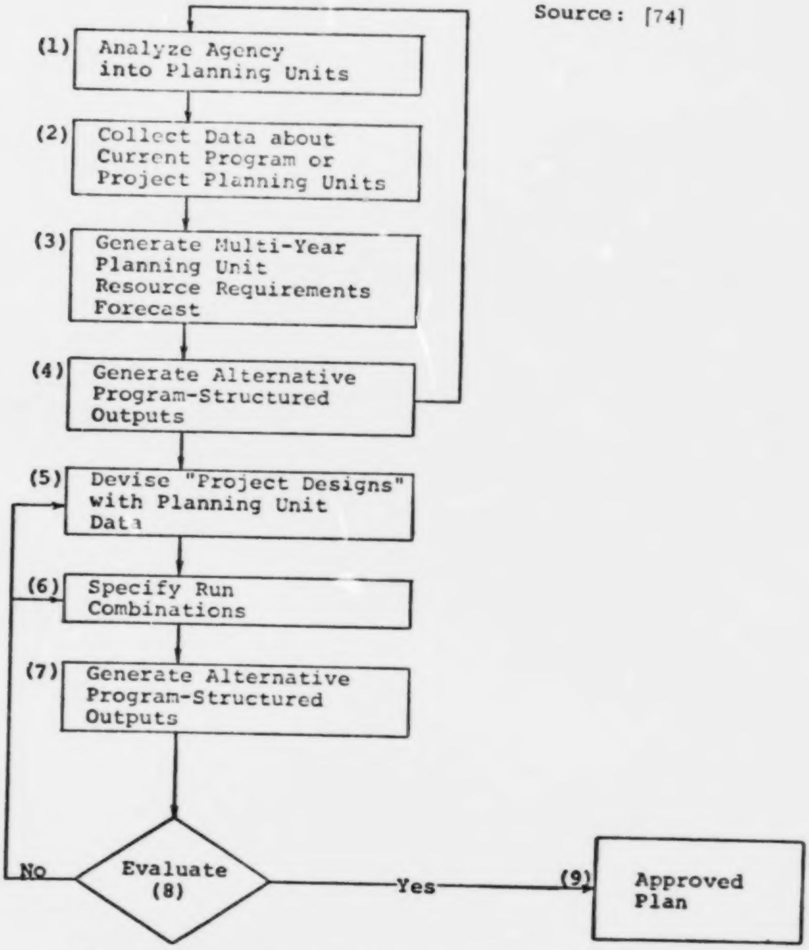
4. Deduce support and indirect operating costs, by applying "planning factors" to the calculation of direct service needs.
5. Using unit costs planning factors, calculate the capital costs, maintenance costs (keyed to sq. ft. of plant), faculty costs, and indirect costs (keyed to faculty costs), and calculate the total costs in each year of the projection.

Clearly, this calculation shows less a prediction that the consequences of certain policy inputs, mainly on hours of facility utilization. By re-iterating the model with several feasible sets of policies, the lowest total cost can be ascertained, and the concomitant policies can be identified as most cost-effective.

In order to develop systematic contingencies on these policies, the model should be run for institutions with different characteristics, to ascertain whether policies should be adjusted for different campuses within a heterogeneous system of schools.

A Model for School District Resource Requirements Planning

The Trenton Public Schools, in New Jersey, has developed a set of models as part of an overall school district planning system [74]. Included are an enrollment forecasting model, a revenue forecasting simulator, and a resource requirements (cost) forecaster (STEP-RRM). In the STEP system, the school district is first divided into "planning units," the smallest division of activities used in planning. A plan is considered a set of planning units, actual or proposed. Alternative plans are produced by adding, deleting, or replacing planning units. The resource requirements model is used to calculate the five-year cost consequences of alternate plans. Figure 6 shows the overall planning process; the resource prediction occurs at steps 3 and 7.


(1)
Analyze Agency into Planning Units
(2)
Collect Data about Current Program or Project Planning Units
(3)
Generate Multi-Year Planning Unit Resource Requirements Forecast
(4)
Generate Alternative Program-Structured Outputs
(5)
Devise "Project Designs" with Planning Unit Data
(6)
Specify Run Combinations
(7)
Generate Alternative Program-Structured Outputs
No
Evaluate (8)
Yes
(9)
Approved Plan

Source: [74]

FIGURE 6—*Resource Requirements Analysis Using the "Planning Unit" Concept*
Source: [74]

The model provides four alternative resource requirements reports: The Planning Unit Report; the Program Report; and Project Report (special clusters of Planning Units); and the Site Report (costs by Facility). Each of these reports can be generated for current programs and projects (known as the "base case" plan) or for planned alternatives. Each report includes:

(1) The number of positions, by each of fifteen staff types, for Year 1-Year 5. (Staff types are re-defined.)

(2) The salary cost, fringe benefit cost, and total, for Year 1-Year 5.

(3) The total capital outlay cost, Year 1-Year 5.

(4) The total non-staff/non-capital outlay cost, Year 1-Year 5.

(5) The total gross cost, and the total local cost.

(6) The total expected positions, by staff type, and total "hires," Year 1-Year 5.

(7) Subsidiary data on planning factors.

These data elements are relatively constant in each of the reports generated by the module; the differences are in the level or focus of aggregation. In the Planning Unit Report, the above information is displayed for each of the district's "planning units." In the Program Report, the information is aggregated in "programs," etc.

STEP-RRM operates by taking detailed, current year data about the "planning units" in the district, along with other inputs about staff unit costs, student/staff ratios, and turnover in the district. Using these and a combination of forecast options available to the planners, the model projects the five-year staff requirements and costs for each planning unit. These planning units are, then, aggregated into large clusters, e.g., programs, projects, sites, or the whole district, to produce the main planning reports.

For each planning unit, users input identification data and information about the current staff, capital outlay, and other resources being utilized in the unit. In addition, appropriate enrollment information (taken from the enrollment forecasting procedure) is input to the planning unit description, along with estimates of categorical or "project" monies expected to accrue to the unit. After data is collected for each planning unit, district data is added: mainly, the "fringe benefit percentage" associated with each staff type, the expected turnover rates associated with each type, and the forecast options chosen for each resource type in each planning unit. The resource requirements model then proceeds to incorporate the effects of projected enrollment change, inflation, turnover rates, and other relevant variables, according to the forecast options selected by the users, and produces cost/resource requirements projections for Year 1-Year 5.

The STEP-RRM, because of its flexibly defined "planning unit" concept, is argued by its developers to be very adaptable to a wide range of school districts and other public service agencies, in both projection of unmet operational costs, or calculating the costs of alternative plans. An interesting feature of the model is that it may be used with any form of program-structure, or several simultaneous program structures. Unlike most PPB-type cost forecasting systems, the STEP-RRM allows for alternative cost analyses without extensive reprogramming or recoding. It is *not* designed for use as a month-to-month program accounting system, but rather as a long-range cost forecaster.

A Model for School Bus Scheduling

The criteria for satisfactory pupil transportation plans have escalated in recent years, at about the same rate as technology for vehicle scheduling has developed. While, on the surface, the task of transporting thousands of students over hundreds of vehicle miles—twice a day—may be considered sufficiently complicated in itself, the problem is made more difficult when we determine to minimize the cost of the operation, or the traveling time for students, or when the objective is tied to a desegregation program.

The school bus scheduling problem has many aspects which cause it to resemble a mathematical programming problem or a "traveling salesman" problem. The unappropriateness of such approaches, however, is apparent when one considers the numbers of variables that would be needed in such models (an objective function with hundreds or thousands of variables, one for each student, for example).

Angel, Noonan, and Whinston [2] have developed and implemented a heuristic model which attempts to:

—minimize number of bus routes
—minimize vehicle mileage
—avert overloading capacity constraints on vehicles
—keep the route time below a maximum determined by policy

The approach developed does not necessarily produce an optimal solution, but it does produce alternative schedules which satisfy the constraints of the problem, from which the schedulers may choose on both quantifiable and non-quantifiable criteria. The algorithm is heuristic, in that it allows users to iteratively develop schedules which are shown to be better or not better than some given schedule; the process is repeated until users are satisfied that any additional runs will not be better then the "best" they have so far.

The scheduling system has three major phases:

1. A data assemblage. Student census, stop locations (on a map grid), and distance data is input,
2. Using the data above, students are assigned to stops, and a set of stop loading cards and a time matrix are generated (the matrix is produced by a modified Moore algorithm, a generalized algorithm for calculating distances in networks), and
3. The stop loadings and time matrix are then augmented with capacity constraint data about number of vehicles, capacity of vehicles, and policy limits on travel time, out of which a solution schedule is generated.

The third part of the process is the most technically interesting and unique. The steps (shown in Figure 7) are:

1. The bus availability table, time/distance matrix, and bus stop loading table are read in.
2. Each stop is assigned to an individual route by entering the bus assigned, the initial stop, the number of students and the total route time into a route table. Stops on a route are kept in a forwardly ordered list.
3. The system calculates a measure of closeness or "association" for each pair of assigned routes, according to several weighted factors (e.g., nearest points, distance from school, current loads, remaining time until school, etc.).
4. Routes are merged as a function of this association measure. The ordering of stops on a given merged route is determined by a traveling salesman algorithm [52], and the process is continued until the number of routes does not continue to decrease. The constraints, of course, are the capacity of the vehicles and the maximum allowable route time.

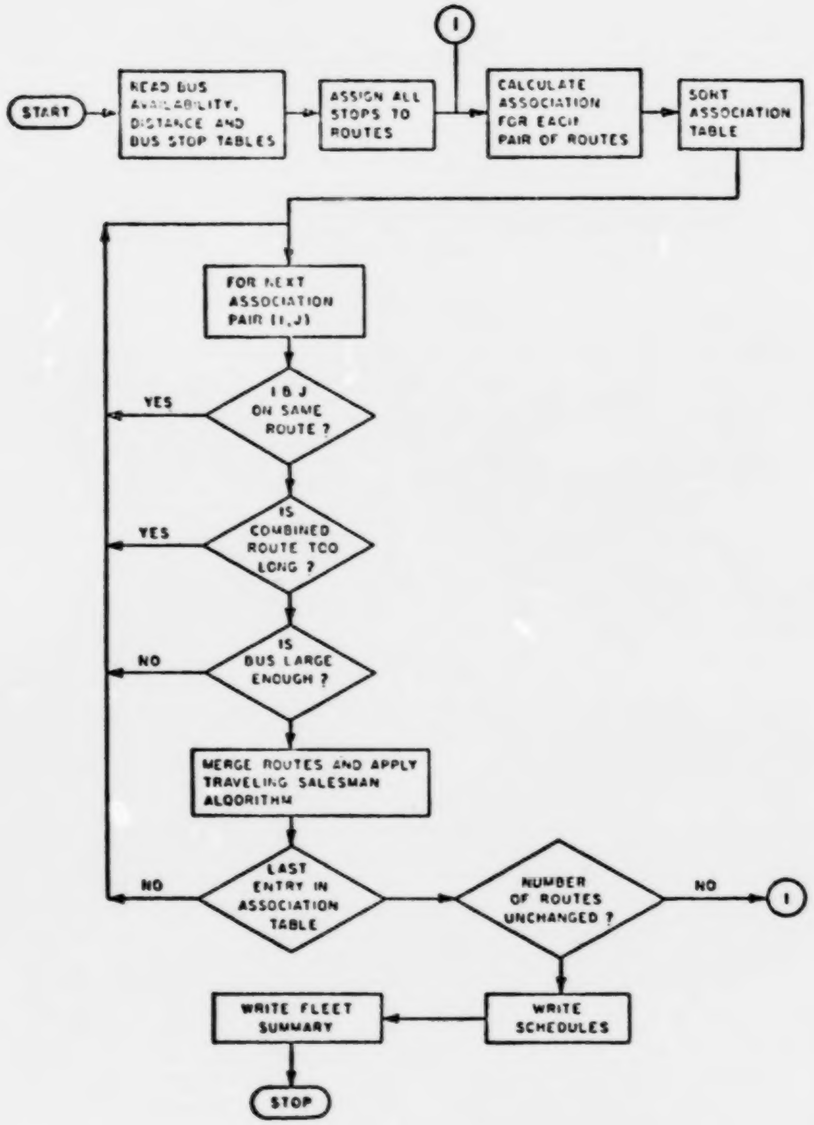
In addition, the authors indicate that the plan shown in Table 2 reduces by one the number of routes in the district and equalizes somewhat the burden on all vehicles (factors which could, presumably, offset entirely the cost of the scheduling job). The system was run on a Control Data 6500 and, as currently reported, can handle problems with as many as 200 grid points.* The developers of the system foresee extensions of the system into long-range planning for fleet requirements.

Table 1 shows a sample model output describing a single route for a given vehicle.

The output includes the location of

---

* Larger districts could use the model by partitioning the community into areas with 150–200 grid points.


START
READ BUS AVAILABILITY, DISTANCE AND BUS STOP TABLES
ASSIGN ALL STOPS TO ROUTES
1
CALCULATE ASSOCIATION FOR EACH PAIR OF ROUTES
SORT ASSOCIATION TABLE
FOR NEXT ASSOCIATION PAIR (I,J)
I & J ON SAME ROUTE ?
YES
IS COMBINED ROUTE TOO LONG ?
YES
IS BUS LARGE ENOUGH ?
NO
MERGE ROUTES AND APPLY TRAVELING SALESMAN ALGORITHM
LAST ENTRY IN ASSOCIATION TABLE
NO
NUMBER OF ROUTES UNCHANGED ?
NO
1
WRITE SCHEDULES
WRITE FLEET SUMMARY
STOP

Source: [2]

FIGURE 7—*Flowchart for Scheduling Sequence*

Table 1

School Bus Route Generation—Southwestern Route Number 3

Stop Coordinates		Stop Numbers		Number of Students	Arrival Times	Load Time
		Old	New			
950*S*	222*W*	129	1	2	4.30	.12
925*S*	135*W*	121	2	2	7.24	.12
1100*S*	200*W*	128	3	9	13.48	.54
1100*S*	100*W*	108	4	2	17.06	.12
1070*S*	0*W*	833	5	5	20.24	.30
1050*S*	55*E*	829	6	3	22.12	.18
1035*S*	100*E*	1301	7	1	23.54	.06
900*S*	200*E*	1340	8	7	29.36	.42
1000*S*	154*E*	1302	9	1	33.48	.06
1050*S*	100*E*	804	10	5	36.24	.30
1100*S*	100*E*	830	11	2	38.06	.12
1100*S*	50*E*	831	12	13	39.30	1.18
1100*S*	10*E*	832	13	10	41.48	1.00
1030*S*	0*W*	816	14	2	44.42	.12
940*S*	0*W*	817	15	2	47.06	.12
925*S*	72*W*	119	16	4	49.24	.24
860*S*	100*W*	120	17	1	52.00	.06

Bus Capacity	72	Total Students	71
Route Time	56.18	Total Stops	17
Loading Time	7.06	Miles Driven	20.5
Driving Speed	25.0	Students/Mile	3.5
Average Speed	21.8	Student-Miles	1324.1

Source: [2]

the stops, in grid coordinates, the number of students at the stop, the arrival time (minutes after the start of the route) and the load time. The report also analyzes the salient characteristics of the route, including total route time and vehicle miles.

Table 2 shows the summary characteristics of the twenty-three routes in a given plan for the Tippencanoe School Corporation.

In the case illustrated, the policy constraints are 72 students per vehicle and 70 minutes plus return maximum travel time; the plan shown above clearly satisfies the constraints.

Importantly, the principal motivation for this project was to improve student *safety*, a concern shared by all managers of school transportation services.

A Model for Class Scheduling

Many colleges and secondary schools have begun to automate the process of assigning students to schedules. While it may not be the soundest educational policy, the typical assignment problem begins with a *fixed* master schedule, to which students must adjust their course preferences. Generally, the only impact of student demand on the master schedule is that undersubscribed courses or sections will be deleted and, in some cases, if the total set of student requests has been tallied in advance, some courses may have sections added to the master schedule.

Macon and Walker [46] point out that most automations of class scheduling behave much in the way that nonautomated systems behave. Instead of a queue of students, the automated system has a queue of student request cards, with sections assigned on a first-come-first-served basis, so that unsatisfied student requests increase as the scheduling proceeds. A slightly more sophisticated automated approach involves the cumulative calculation of students assigned, so that requests for multiple-section courses can be routed to the least subscribed sections, thereby reducing the number of student dissatisfactions. Nevertheless, this approach is still, es-

Table 2

Summary of Schedule

Number of routes: 23
Average load size: 66
Load size—range: 57–72
Maximum riding time: 58 minutes

Route Number	Scheduled Quantity	Number of Stops	Number of Miles	Max. Riding Time (min.)
1	67	32	13	43
2	65	26	13	38
3	72	37	12	50
4	69	38	14	52
5	72	20	13	31
6	71	33	12	48
7	65	30	8	40
8	71	26	11	40
9	65	33	12	46
10	70	26	10	37
11	61	24	11	33
12	72	31	25	47
13	64	13	15	25
14	66	32	15	25
15	70	46	7	41
16	62	30	8	40
17	63	33	24	44
18	60	27	28	44
19	57	22	12	31
20	70	21	13	34
21	70	37	25	47
22	61	29	21	45
23	57	30	17	47

Source: [2]

sentially, a first-come-first-served approach which unjustifiably penalizes students who happen to come later in the queue.

Macon and Walker propose a "fairer" alternative to scheduling, using a "Monte Carlo" model. They posit the most difficult case, namely, that in which students give only "first choices" (the case is more typical in colleges than in secondary schools), and consider any requested student schedule which cannot be assigned as "unable to be processed." Their goal, thus, is to minimize the number of unprocessable requests.

To accomplish this goal, the total number of requests for each section is tallied in advance, and for each section one computes a value p defined as

$$p = \min\left(1, \frac{\text{no. of seats available}}{\text{no. of requests remaining}}\right)$$

(i.e., minimum of 1 and the ratio)

In the first iteration, of course, the number of requests remaining is the total unprocessed set of student requests. Given this index . . .

> "processing a student's list of sections desired proceeds in the following way: A random r between 0 and 1 is compared to the p-value of the first section requested. If $r \leq p$ the section is assigned and the remaining values for the sections available and requested are modified accordingly. If $r > p$ the other sections of the same course are tested one at a time for conflicts with sections already assigned to the student and the requests remaining on his list. If no conflicts occur and the corresponding p-value is greater than that associated with the section requested, then the section requested is replaced by the section tested. This continues until all sections of the course have been similarly examined. The conflict-free section with the largest p-value is thus assigned. . . ."

Inevitably, if any section of any course is closed, there is a possibility of being unable to process a request. There is the special case, however, in which the assignment of students was based on preferences, while another section assignment would still have been feasible (not conflicting with other requests by the same student). Macon and Walker have also developed an assignment program which searches the individual student assignments to see if sections can be changed for individual students without conflict, thereby opening some sections which were previously closed. The assignment model increases the probability of being able to process all requests, and also distributes section enrollments more easily.

Using an actual master schedule of over 2000 sections, the users were able to assign 9000 students (averaging 4 section requests per list) in under four minutes of machine time. The authors believe that with a few passes of the model, the number of unprocessable student requests will be reduced to 0, or to so few that they may be solved with special attention.

It should be reinforced that this, and other sophisticated scheduling systems, assume a given master schedule which is relatively inflexible. As indicated earlier in this chapter, this approach satisfies administrative needs but may be inconsistent with a more sophisticated approach to determining what courses should be offered in the first place.

A Theoretical Model for Instructional Optimization

As indicated earlier in this chapter, the most pervasive, and perhaps important, educational decisions involve the minute-to-minute assignments of students to instructional tasks. But, as was also indicated, this process goes on almost independent of any formal, let alone mathematical, model of instruction. Of course, as Sisson observed some years ago, what takes place in the classroom decision-making is, in a real sense, influenced by many models "in the head" of instructors [61].

The discipline of curriculum engineering is concerned with formalizing instructional models and developing alternatives to the classroom teacher as the decision-maker (alternatives such as the student, a machine, a non-professional monitor). The measures of instructional effectiveness which are to be improved include student performance (higher probability of achieving instructional objectives), time (reduced time), cost, or motivation (by finding the optimum "style" or "strategy" for a given student). The common attribute of the diverse approaches to curriculum engineering and instructional technology is a concern with defining a terminal objective for the instructional program, defining a set of intermediate objectives en route to the terminal objective, and specifying all these objectives in a form which allows accurate appraisal of the student's progress, to be used as the basis for deciding what instructional task to be assigned to the student at any given decision point. There are, in such systems, devices for providing instruction (people, machines, material) and devices for monitoring progress (people, machines, material). In the technique known as CAI (computer-assisted-instruction), the computer is both the provider and the monitor. In CMI (computer-mediated, or computer-managed instruction) the computer is the monitor which evaluates data from or about students and makes an instructional assignment which is carried out by other machines, material, or people.

The generalized approach to instruction assumed here is most familiar in "programmed instruction" and the "programmed text"; this device, frequently attributed to B. F. Skinner, makes it impossible for a student to move through a book without periodically demonstrating mastery of an intermediate objective (an attribute of all engineered curricula). Programmed instructional materials are generally designed on a linear (not in the sense of linear programming) or branching basis, as shown in Figure 8.

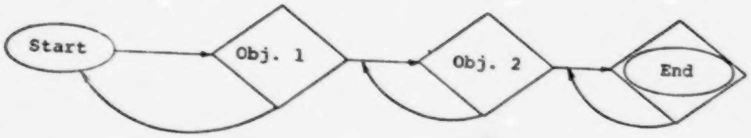
As this representation shows, the most typical arrangements allow for, either:

a) the student repeats a task until he can satisfy the criterion for the intermediate objective, after which he continues. (linear)

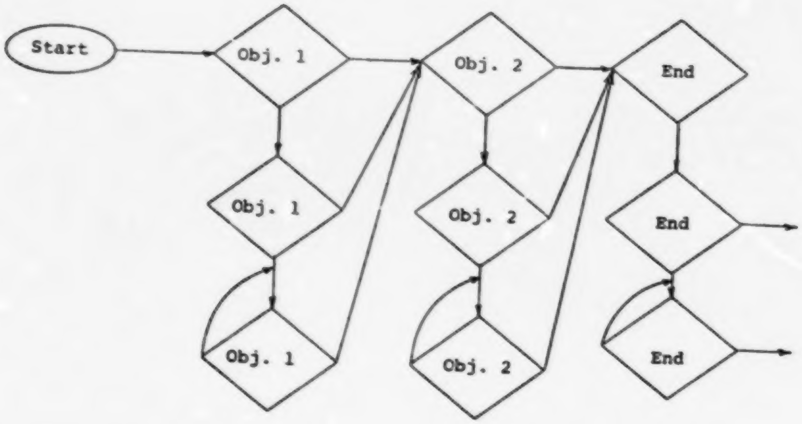
b) the student who fails a task is routed to another path, another task which is remedial or alternative in style; he can be re-routed indefinitely, depending on the complexity of the program design, until, finally, he begins to loop around one task (or is randomly assigned to another, or terminated as student, etc.). Thus there are multiple paths through the process. (branching)

Clearly, branching instructional programs can be rendered as a matrix, and manipulated algebraically. Belgard and Min [9] depict this array as shown in Figure 9.

As shown in the illustration, a student moves through a series of stages. The "linear" path is $\mathbf{K}_{11}$–$\mathbf{K}_{1n}$, the initial line; however, a student who fails to


Start
Obj. 1
Obj. 2
End

Linear Instructional Program


Start
Obj. 1
Obj. 1
Obj. 1
Obj. 2
Obj. 2
Obj. 2
End
End
End

Branching Instructional Program

FIGURE 8—*Design of Programmed Instructional Materials*

j = 1, …, n i = 1, …, m			STAGE					
			Task (1)	Task (2)	Task (3)	…	Task (n)	
	Initial	1	K_{11}	K_{12}	K_{13}	…	K_{1n}	$K_{1\cdot}$
S	Alternative	2	K_{21}	K_{22}	K_{23}	…	K_{2n}	$K_{2\cdot}$
T		3	K_{31}	K_{32}	K_{33}	…	K_{3n}	$K_{3\cdot}$
A		4	K_{41}	K_{42}	K_{43}	…	K_{4n}	$K_{4\cdot}$
T								·
E						K_{ij}		
		m	K_{m1}	K_{m2}	K_{m3}	…	K_{mn}	$K_{m\cdot}$
			K.1	K.2	K.3	…	K.n	

Source: [9]

FIGURE 9—*Matrix Representation of Branching Instructional Program*

move rightward on the initial path may branch to another task at the same stage. Actually, this matrix representation is more general than a typical branching program. For purposes of instructional optimization, *every permutation of tasks* may be considered a path to the terminal instructional objective. Each path has a differential probability of attaining the terminal objective, and each path has different costs (material and non-material). Of course, several of the possible permutations are highly improbable, but allowing the broadest range of alternatives also allows the broadest possible differentiation of students, in terms of psychological needs and "learning styles."

The Belgard-Min model is based on *probabilistic* concepts of the teaching learning process. Thus, at each stage in the matrix, the student performance is measured, and the probabilities are calculated for every candidate next stage. Optimality is a function of maximizing probability of attaining the terminal objective (assuming that it can be estimated) within physically and psychologically defined constraints, and it is argued that global optimality for any student is the sum of next-task optimalities (the State-Increment Model).

The further mathematical details of the model are less important here than an appraisal of the general approach. As already mentioned, there are curricula which are, in fact, constructed as a matrix or network through a set of measurable objectives, and there are automated systems which monitor and/or provide instruction in these curricula. The limitation on such a concept is, clearly, that many educators believe they object to this conception of instruction, allowing that it may be suitable for "training," or even "skills instruction," but not for "education." It is difficult to say where the burden of proof lies in such a debate, but it is clear that there will be increasingly larger branches of instruction which lend themselves to this approach. It also appears that implementation will occur more readily in military, industrial, vocational, and profit-oriented instruction, and that the public schools and colleges will be among the last to make significant moves in this direction.

III. AN ASSESSMENT OF DECISION MODELS IN EDUCATION

Introduction: Assessment Criteria

Models in educational planning and operation exist to serve a wide variety of purposes—particularly in the broadly defined area of planning. For that reason, assessments should be function-specific, a review of the adequacy and availability of models devised for each specific purpose.

In the next section we shall attempt to assess the models available to perform various functions for various levels of the education complex. Some *rating*, therefore, will be involved, based on the following scale, from highest to lowest:

—Best:	Models are available, accessible, and in limited use. (4)
—	Models are available and accessible, but little used. (3)
—	Models are available, but largely inaccessible and in little use at this time. (2)
—Worst:	Models are of limited availability, or virtually non-existent. (1)

Models will be appraised according to the following functions:

a. Models for specific short-run operational decisions

b. Models for longer-range forecasting and predictions

c. Models for informing non-quantifiable policy decisions (including some of b.)

d. Models for stimulating research and inquiry related to decision problems (including some from all the above)

Below, we shall indicate judgments about the rating of models, for each function, at the key decision levels mentioned in the early part of the chapter. Finally, we shall speculate about developments in the near future, including factors which will impede or facilitate further development.

Assessment

Table 3 shows the author's assessment of current models in educational planning and operation. While it would be improper to attach too much significance to these ratings, the table does indicate some interesting general trends:

1. The level for which models seem to be most available, accessible, and used, is the national level (except in the area of short-run operational decisions, which is not applicable). The national level makes relatively heavy use of forecasting systems, and projection models are in many respects easier to build for large, gross forecasts than small, detailed ones. Several units of the Division of Education are concerned with modeling, particularly forecasting, and it is not surprising that this relatively high level of development and utilization exists. (The use is not universal, but intensive in certain units.)
2. Forecasting and Prediction models are the functional area with greatest development and utilization at all levels, approximately equalled by short-run operating decision models, which, when applicable, seem to be highly developed.
3. Forecasting and prediction models are more likely to be used strictly for projections than in a simulation mode, that is, to project the consequences of alternative policy decisions. Thus, they are infrequently used in the function of informing non-quantifiable policy decisions.
4. One of the most advanced areas of model development is in the variety of detailed operating decisions made at the school, school district, college, or college system level. These include a variety of approaches to scheduling, inventory, staff assignment, vehicle deployment, etc. Interestingly, the advances in these areas may be due to the wide availability of comparable models already developed to solve analogous prob-

Table 3 Assessment of Education Models

Level	Function			
	Short-Run Op. Decisions	Forecasting Prediction	Policy Informing	Stimulating Research
National: Policy/Program	n/a	4	3	3
State: Subsidy/Finance	3	3	2	2
State: Staff	2	2	2	2
State: School Approval	n/a	1	2	2
State: Boundaries/Jurisdiction	3	4	2	2
School District: Student Decisions	3	3	2	3
School District: Operating Decisions	4	3	2	2
School District: Program/Policy	n/a	2	2	1
College: Student Decisions	4	4	3	3
College: Operating Decisions	4	4	3	3
College: Program/Policy	n/a	4	3	2

Ratings: 4 = available, accessible, limited use
3 = available, accessible, little use
2 = available, largely inaccessible, little use
1 = limited availability or non-existent

lems in business and industry. Also, one should attribute the prevalence of modelling in these areas to the activities of disseminating agencies such as the National Center for Higher Education Management, the Center for the Advanced Study of Educational Administration, the National Center for Educational Statistics, private consulting firms, and other organizations dedicated, in part, to the extended utilization of quantification in educational decision-making.

The Future of Models

Whether there will be extended use of models in the future depends on at least the following four factors.

First, modelling as a science and craft poses its own theoretical and technical problems, which frequently must be solved before the model can be used with a level of effort and cost proportionate to its value. The increased use of linear programming in recent years can be partially attributed to the development of solution algorithms and efficient computer procedures which make linear programming feasible to use. At present, for example, many simulation models used for resource requirements projection require enormous machine capacity and cost. More efficient approaches are needed.

Second, there are many theoretical problems inherent to education itself which limit the possibility (or, at least, the precision) of models. We still lack important theoretical insights into the teaching-learning process, for example, and can construct instructional decision models based only on the most superficial measures, very few of them related to the important "soft" outputs of education [5, 11, 12, 13, 17, 18].

Third, there is the need to appropriately document, publish, package, and market modeling materials, to close the time gap between development and implementation. The key marketing devise, here, is training of potential users in the value and technique of modeling, assuring that in each candidate agency there is at least one manager who can identify a decision which will be served by models and can communicate his needs to appropriate technical staff or consultants.

Fourth, and finally, the barrier of emotional resistance to quantified methods must be overcome. This resistance in education comes from two main sources: (a) politically oriented educational leaders who prefer politically attractive solutions to technically justified ones, and (b) educators who, for good or bad reasons, believe that an emphasis on quantification and formalization of educational decision processes is inconsistent with the well-being of students.

The first and third of the factors (better modeling technique and marketing) are not severe impediments; both can be ameliorated by the application of more resources to the area. The second and fourth (educational theory and emotional resistance) are vastly more difficult to overcome or skirt. To have an impact on these two barriers will require many years of effort by modelling advocates; it will require better models, better packaged; and it will also probably require appropriate changes in the professional training given to educators.

References and Selected Bibliography

1. Alkin, M. C., "The Use of Quantitative Methods as an Aid to Decision Making in Education," presented to American Educational Research Association (Feb., 1969).
2. Angel, R. D., Caudle, R. N., and Whinston, A., "Computer-Assisted School Bus Scheduling," *Management Science, 18,* No. 6, pp B–279–B–288 (1972).
3. Armitage, P. M., Smith, C. S., and Alper, P., *Decision Models for Educational Planning*, Penguin, London (1969).
4. Averch, H., et al., *How Effective is Schooling? A Critical Review and Synthesis of Research Findings*, Rand Corporation, R–956–PCSF/RC (1972).
5. Averch, H. and Kiesling, H., "The Relationship of School and Environment to Student Performance: Some Simul

taneous Models for the Project TALENT High Schools," Rand Corporation (mimeo, 1970).

6. Balinsky, W. and Reisman, A., "A Taxonomy of Manpower-Educational Planning Models," *Socio-Economic Planning Sciences, 7*, No. 1, pp 13–18 (1973).
7. Banghart, F., *Educational Systems Analysis*, Macmillan, New York (1969).
8. Bani, K., Gupta, S., and Sisson, R., "Toward Aggregate Models of Educational Systems," *Socio-Economic Planning Sciences, 3*, No. 1, pp 25–36 (1969).
9. Belgard, M. R. and Min, L. Y., "Optimizing the Teaching-Learning Process Through a Linear Programming Model—Stage Increment Model," presented at American Educational Research Association (February, 1971).
10. Benard, J., "General Optimization Model for the Economy and Education," in *Mathematical Models in Educational Planning*, Organization for Economic Cooperation and Development (OECD), Paris (1967).
11. Bowles, S. and Levin, H., "More on Multi-Collinearity and the Effectiveness of the Schools," *Journal of Human Resources, 3*, No. 3, pp 393–400 (1968).
12. Bowles, S., "The Determinants of Scholastic Achievement—An Appraisal of Some Recent Evidence," *Journal of Human Resources, 3*, No. 1, pp 2–24 (1968).
13. Bowles, S. S., "Towards an Educational Production Function," in W. L. Hansen (Ed.), *Education, Income, and Human Capital*, Columbia University Press, New York (1970).
14. Bruno, J., "A Mathematical Programming Approach to School System Finance," *Socio-Economic Planning Sciences, 3*, No. 1, pp 1–12 (1969).
15. Bruno, J. E., "A Proposed Scheme for Federal Support for Education," Rand Corporation, P-4038 (1969).
16. Bruno, J. E., "Using Linear Programming Salary Evaluation Models in Collective Bargaining Negotiations With Teacher Unions," *Socio-Economic Planning Sciences, 3*, No. 1, pp 1–12 (1969).
17. Bureau of Educational Personnel Development, Office of Education, *How Do Teachers Make a Difference*, HEW, OE-58042 (1970).
18. Burkhead, J., Fox, T. G., and Holland, J., *Input and Output in Large City High Schools*, Syracuse University Press, Syracuse (1967).
19. Centner, S. I., Judy, R. W., and Levine, J. B., *Campus V: Documentation*, University of Toronto, Institute for the Quantitative Analysis of Social and Economic Policy (January 1970).
20. Clarke, S. and Surkis, J., "An Operations Research Approach to Racial Desegregation of School Systems," *Socio-Economic Planning Sciences, 1*, No. 2 (1968).
21. Cohn, E., "Economies of Scale in Iowa High School Operations," *Journal of Human Resources, 3*, No. 4, pp 422–434 (1968).
22. Copa, George M., "Identifying Inputs Toward Production Function Applications in Education," Research Coordinating Unit for Vocational Education, University of Minnesota (1971).
23. Correa, H., *Models and Mathematics in Educational Planning*, World Yearbook of Education, pp 398–442 (1967).
24. Csima, J. and Gotlieb, C. C., "Tests on a Computer Method for Constructing School Timetables," *Communications of the ACM, 7*, No. 3, pp 160–163 (1964).
25. Fox, K. A., *Formulation of Management Science Models For Selected Problems of College Administration*, Final Report for HEW Contract OEC-3-068058, Iowa State University (1967).
26. Froomkin, J., "Cost/Effectiveness and Cost/Benefit Analysis of Educational Programs," *Socio-Economic Planning Sciences, 2*, No. 2, pp 381–388 (1969).
27. Gerwin, D., "A Process Model of Budgeting in a Public School System," *Management Science, 15*, No. 7, pp 338–361 (1969).
28. Graves, R. J. and Thomas, W. H., "A Classroom Location-Allocation Model for Campus Planning," *Socio-Economic Planning Sciences, 5*, No. 2, pp 191–204 (1971).
29. Hammond, A. Y., "Mathematical Models in Education and Training," Rand Corporation, RM-6357-PR (1970).
30. Harden, W. R. and Tcheng, M. T., "Classroom Utilization by Linear Programming," Office of Academic Planning, Illinois State University (1971).
31. Heckman, L. B. and Taylor, H. M., "School Rezoning to Achieve Racial Balance: A Linear Programming Approach," *Socio-Economic Planning Sciences, 3*, No. 2, pp 127–133 (1969).
32. Holtzman, A. G., "Linear Programming and the Value of an Input to a Local Public School System," *Public Finance, 23*, No. 3, pp 429–440 (1968).
33. Holtzman, A. G. and Turkes, W. R. (Eds.), *Optimal Scheduling in Educational Institutions*, U.S. Office of Education, Cooperative Research Project 1323, University of Pittsburgh (1964).
34. Hopkins, D., "On the Use of Large-Scale Simulation Models for University Planning," *Review of Educational Research, 41*, No. 3, pp 467–478 (1971).
35. Hussain, K. M., *A Resource Requirements Prediction Model (RRPM-1)—Guide for the Project Manager*, National Center for Higher Education Management Systems at WICHE, Technical Report 20 (October 1971).
36. International Business Machines, *EPIC: Budget and Finance—General Information Guide*, IBM, GH20-1132-0 (1972).
37. Jaffee, A. J., *Handbook of Statistical Procedures for Long Range Projections of Public School Enrollment*, Office of Education, OE-24017 (1969).

38. Jewett, J. E., "College Admissions Planning: Use of a Student Segmentation Model," Office of the Vice President—Planning, University of California (November, 1971).

39. Karraker, R. J. and Dolce, L. A., "Errorless Discrimination of Alphabet Letters: Effect of Time and Method of Introducing Competing Stimuli," *Journal of Experimental Education, 38*, No. 1, pp 27–35 (1970).

40. Hu, T., Lee, M. W., and Stromsdorfer, E. W., *A Cost-Effectiveness Study of Vocational Education*, Institute for Research on Human Resources, Pennsylvania State University (March 1969).

41. Kiesling, H. J., *Input and Output in Compensatory Education Projects in California*, Rand Corporation, R-781-CC/RC (October, 1971).

42. Kiesling, H. J., *The Relationship of School Inputs to Public School Performance in New York State*, Rand Corporation, P-4211 (October 1969).

43. Knezevich, S. J., *Administrative Technology and The School Executive*, American Association of School Administrators, Washington, D.C. (1969).

44. Koenigsberg, E., "Mathematical Analysis Applied to School Attendance Areas," *Socio-Economic Planning Sciences, 1*, No. 4, pp 465–475 (1968).

45. Levin, H. M., "Concepts of Economic Efficiency and Educational Production," NBER Conference on Education as an Industry, June 4–5, 1971.

46. Macon, N. and Walker, E. E., "A Monte Carlo Algorithm for Assigning Students to Classes," *Communications of the ACM, 9*, No. 5, pp 339–340 (1966).

47. Mathematica, Inc., *Enrollment and Financial Aid Models for Higher Education*, Bethesda, Maryland (August, 1971).

48. McNamara, J., *A Mathematical Programming Model for the Effective Allocation of Vocational Technical Education Funds*, Pennsylvania Department of Education (1970).

49. McNamara, J., "A Regional Planning Model for Occupational Education," *Socio-Economic Planning Sciences, 5*, No. 3, pp 317–399 (1971).

50. McNamara, J., "Mathematical Programming Applications in Education," *Socio-Economic Planning Sciences, 7*, No. 1, pp 19–36 (1973).

51. National Educational Finance Project, *Economic Factors Affecting the Financing of Education (Vol. 2)*, NEFP, Washington (1970).

52. Newton, R. and Thomas, W. H., "Design of School Bus Routes by Computer," *Socio-Economic Planning Sciences, 3*, No. 1, pp 75–85 (1969).

53. Oliver, R. M., "Models for Predicting Gross Enrollments at the University of California," Office of the Vice President—Planning and Analysis, University of California (August, 1968).

54. Organization for Economic Cooperation and Development, "A Simulation Model of the Educational System," OECD, Paris (1970).

55. Organization for Economic Cooperation and Development, *Econometric Models in Education: Some Applications*, OECD, Paris (1965).

56. Organization for Economic Cooperation and Development, *Mathematical Models in Educational Planning*, OECD, Paris (1967).

57. Pfeiffer, J., *A New Look at Education—Systems Analysis in Our Schools and Colleges*, Odyssey Press, New York (1968).

58 Piland, J. C. and Lemke, E. A., "The Effect of Ability Grouping on Concept Learning," *Journal of Educational Research, 64*, No. 2, pp. 209–212 (1971).

59. Ploughman, T., "An Assignment Program to Establish School Attendance Boundaries and Forecast Construction Needs," *Socio-Economic Planning Sciences, 1*, No. 2, pp. 243–258 (1968).

60. Sisson, R., Jacobson, L. and Robinson, S., "The CCHE Facilities Analysis Model," MATHEMATICA, Inc., Princeton (October, 1970).

61. Sisson, R., "Can We Model the Educational Process," *Socio-Economic Planning Sciences, 2*, No. 2, pp. 109–120 (April 1960).

62. Stoller, D. and Dorfman, W. (Eds.), "Proceedings of the Symposium on Operations Analysis of Education," *Socio-Economic Planning Sciences, 2*, Nos. 2, 3, 4, pp. 103–520 (1969).

63. Taft, M. I. and Reisman, A., "Toward Better Curricula Through Computer Selected Sequencing of Subject Matter," *Management Science, 13*, pp. 448-468 (1967).

64. Temkin, S., *Benefit-Cost Analysis and Effectiveness-Cost Analysis: A Prospective for School Administrators*, Research For Better Schools, Philadelphia (1969).

65. Thomas, J. A., *The Productive School: A Systems Approach to Educational Administration*, Wiley, New York (1971).

66. Turksen, I. B. and Porter, E. D., "Micro Level Forecasting of Student Enrollment with an Information System," Department of Industrial Engineering, University of Toronto (n.d.).

67. Turksen, I. B. and Holzman, A. G., "Short Range Planning for Educational Management," presented to Operations Research Society of America (October, 1970).

68. University of the State of New York, "New York State Performance Indicators in Education," USNY, Bureau of School Programs Evaluation, Albany (1972).

69. Van Dusseldorp, R., Richardson, D., and Foley, W. *Educational Decision Making Through Operations Research*, Allyn and Bacon, Boston (1971).

70. Wasik, J .L., "The Development of a Mathematical Model to Project Enrollments in a Community College System," presented to American Educa-

tional Research Association (February, 1971).

71. Weathersby, G. B. and Weinstein, M. C., "A Structural Comparison of Analytical Models for University Planning," Office of the Vice President—Planning and Analysis, University of California (August, 1970).

72. Weathersby, G., "Educational Planning and Decision-Making: The Use of Decision and Control Analysis," Office of the Vice President—Planning and Analysis, University of California (1969).

73. Weathersby, G. B., "Student Tuition Models in Private and Public Higher Education," presented to Operations Research Society of America (October, 1970).

74. Weiss, E. H. and Ackerman, J., *System for Trenton's Educational Planning (STEP): User's Manual*, Trenton Public Schools (1972).

75. Zemach, R., "A State-Space Model for Resource Allocation in Higher Education," *IEEE Transactions in Systems Science and Cybernetics*, *4*, No. 1, pp. 108–118 (1968).

Chapter 10

Models in the Field of Energy

By

Kenneth W. Webb

SUMMARY 319
I. INTRODUCTION 320
Categories of Decisions 321
II. INDUSTRIAL DECISION MAKING AND THE USE OF ENERGY MODELS 321
Sources of Energy 322
Exploration Decisions 323
Mining Decisions 323
Transportation to Energy Manufacturing Plants 323
Manufacturing and Distribution Decisions 324
III. GOVERNMENT DECISION MAKING AND THE USE OF ENERGY MODELS 325
Projections by Trend Extension 326
A Demand Estimation Model 327
A Dynamic Demand Model With Interfuel Competition 328
A Pollution Policy Model 331
A Model for Rate Making Proceedings 332
A Model for Federal Procurement of Fuels 333
A Model to Locate Needed Research 334
A Model of Oil Supply and Distribution 335
A State Government Demand Model 338
IV. RECENT DEVELOPMENTS AND NEW DIRECTIONS 339
REFERENCES 343

Models in the Field of Energy

SUMMARY

Rising fuel prices, shortages, and the estimates of future shortages, if not depletions, has caused increased concern in the energy industry and Federal government. The procedures of environmental protection have further complicated the energy supply and price picture. Pollution abatement and facility siting problems have delayed construction of facilities, decreased production and increased cost. The question of how serious an energy problem we are going to have and alternative strategies for alleviating the problem can be answered by statistical estimating procedures and more extensive models. The adequacy of the available models and the promises of the models under development are under sharp review.

This chapter is concerned with the decision making that occurs in the field of energy and the models that are used as aids in making those decisions. The Federal government makes use of models in many decision areas such as stimulating exploration, determining rates, conserving energy, import quotas and environmental quality. State and local government are beginning to use modeling for problems of locating facilities, pollution control, safety and others.

Non-governmental institutions have been using models far longer than their government counterparts in such decision making areas as manufacture of oil and gas, financing projects, distribution of products, and market analysis. The first section of this chapter describes the decision making in the industrial world, and the models that are used by industry. The exploration for reserves of oil, natural gas, uranium and coal has increased modeling efforts in terms of statistical estimation of sampled reserves and economic evaluation of the field. Mining operations use PERT and simulation for developing and testing operational designs. The transportation systems from reserve areas to manufacturing sites makes use of scheduling techniques and inventory control methodology. Many linear programming models are used in the manufacturing of energy, especially in the oil industry. Complex models that represent several refineries are in operation as aids to reducing cost and increasing output. Electricity utilities are using network analysis to determine the optimum configuration of plants and transmission systems. In general, the industrial uses of models have been to increase profit or supply more service in the case of utilities. The questions of degradation of the environment and social issues are only recently forcing their way into the industrial decision processes.

The main body of this chapter on energy modeling is on the models of use to the government. The Federal government needs models that will permit the analysis policy affecting demand, supply and cost. The models do not exist today that will answer these needs; however, models are being used and studied which partially respond to the needs. This chapter reviews models that are used in trend projection of exploration, manufacturing, demand, price and other components of the estimating problem. The problems of using trend projection, where there appears to be complete changes in relationships of supply and demand, is discussed in detail.

Preceding page blank

Econometric models that are useful to government planning are also discussed, including a dynamic demand model with interfuel competition, a pollution policy model for testing new government pollution policies, and a linear programming model for the procurement of fuels by the Federal government.

The chapter contains a description of a model useful for determining the areas of energy supply where research monies can best be allocated, a model that was used by the Canadian government for the assessment of the effects of the Arctic oil findings, and a State government model for examining the effects of price changes in power and developed for a twenty year power plant citing plan. Recent econometric models and research efforts are also described to illustrate the complexities that occur in modeling when more comprehensive estimates and models are required. The following summary table contains the main models discussed in this chapter.

I. INTRODUCTION

Energy decision problems are among the most critical ones confronting both Federal and State governments. For the past thirty years, the relative costs of most forms of energy have been decreasing and the availability far exceeded the demands. However, the year 1971 was the first year that the average price of electricity rose; and, in 1972, shortages of natural gas, oil and electricity prevented expansion of usage in certain sections of the country.

Predictions indicate that the supply of energy to meet demands will continue to be a problem and that costs will increase. Economists feel that increasing costs will tend to slow the use of energy and have a depressing effect on the economy.

The problems of supply and cost have been complicated by the recent increased demands for environmental protection. The location of facilities to preserve the environment, the use of

ENERGY

Models Discussed in Chapter

Summary Table

Model Decision Area	General Type	Important Characteristics
Stimulation of Exploration for new sources	Dynamic Model	Combines supply, demand and price with interfuel competition. Assesses the effects of new reserves
Determination of Rates	Regression	Models exploration, residential, commercial and industrial consumption
Conservation of Energy	Dynamic Model	Combines supply, demand and price with interfuel competition
Stimulation of new methods of manufacture	Engineering Equations	System models to determine the benefits of changes in the system
Import Quotas	Network Model	Assesses the effects of reserves and imports
Purchase of products for government consumption	Linear Programming	The least cost selection of energy vendors to fill government needs is determined
Power Plant Siting	Deterministic Scenario Analysis	A State government analysis of demand projections and location effects of facilities
Pollution Abatement	Linear Programming	Subject to sulfur controls the model predicts supply, demand and price by air quality regions

pollution abatement procedures, and general conservation restrictions have added cost and retarded new production of energy.

The continuance of supply problems may require the government to make difficult decisions concerning the conservation of energy. These may include changes in building codes such as increased insulation, tax incentives to use less energy, and advertising the benefits of conservation practices. A more careful use of our energy could ease some environmental problems and lessen the need for imports which impact our balance of payments.

Categories of Decisions

Insuring an adequate supply of energy has become a critical decision making area at all levels of government. Decision making at the Federal level of government is concerned with the following:

- stimulation of exploration for new sources
- determination of rates
- conservation of energy
- stimulation of new methods for manufacturing energy
- import quotas
- direct purchase of products for government
- environmental quality
- assuring the growth of the economy through sufficient supplies of energy.

Decision making at the local government level is concerned with the following:

- location of facilities such as new power plants
- pollution control and other environmental impacts
- safety considerations
- direct purchase of products for local government
- assuring growth of local economies relative to their dependence on energy
- in some localities, the operational problems of operating government power utilities.

In addition to these governmental areas of decision making are the decisions of many institutions concerned with the production and distribution of energy. They include the following:

- companies that manufacture oil, electricity, natural gas and other products
- national agencies that manufacture electricity such as the Tennessee Valley Authority
- commercial banks that finance projects
- the courts that adjudicate the problems of location of facilities, rates and regulation
- scientific groups that are concerned with all aspects of energy supply and use
- environmental and conservation groups that bring suit and use persuasion on problems affecting the environment
- the general public which makes decisions by voicing attitudes and by referendum.

The use of models in decision making in industry is well developed. The use of models by government is in the beginning stages. By way of introduction to models in energy, we shall first briefly review the industrial models to illustrate the extensiveness of their use. Following this, detailed descriptions will be made of governmental decision making relative to the use of models.

II. INDUSTRIAL DECISION MAKING AND THE USE OF ENERGY MODELS

Energy for the use by society is generally supplied by private enterprise. The exceptions are the electricity utilities that are sometimes part of city government or such bodies as the Tennessee Valley Authority. In general, however, the production of energy products such as coal, natural gas, oil and electricity is in the private domain. These products are dependent on basic sources that are found in or on the surface of the earth, as shown in Chart 1.

By far, the largest basic source of energy is solar radiation at the surface of the earth. Solar energy is captured by leaves of plants where it is stored by the process of photo-synthesis. The leaves get buried and eventually provide the fossile fuels. A recent analysis of the possible total of the world's mineable coal indicates 7.6 trillion tons. The current production rate is 3 billion tons each year. Thus, there are two to three centuries of reserve coal if the present production rate does not double more than three times. Oil and natural gas are estimated to be in much shorter supply. For instance, the peak of production of natural gas is estimated to be about 1980.

Heat from the interior of the earth is presently a very minor source of energy. Several plants exist for the production of electricity. One has been in operation in the Larderello part of Italy since 1904 and currently has a capacity of 370 megawatts. The use of tidal and wind power is still experimental. One tidal power plant for the production of electricity is on the Channel Island off the coast of France and produces 240 megawatts.

In the field of nuclear energy, the basic input is Uranium 235. The Atomic Energy Commission estimates that it will take 206,000 short tons of U_3O_8 to meet the needs of the decade 1970 to 1980. Extrapolating trends, the AEC estimates that the quantity of ore producible at $8 per pound in the United States is 243,000 short tons, indicating a need for additional reserves to be found.

The descriptions of the supply picture for several energy products is characterized by diminishing resources. This is going to result in both short supply for

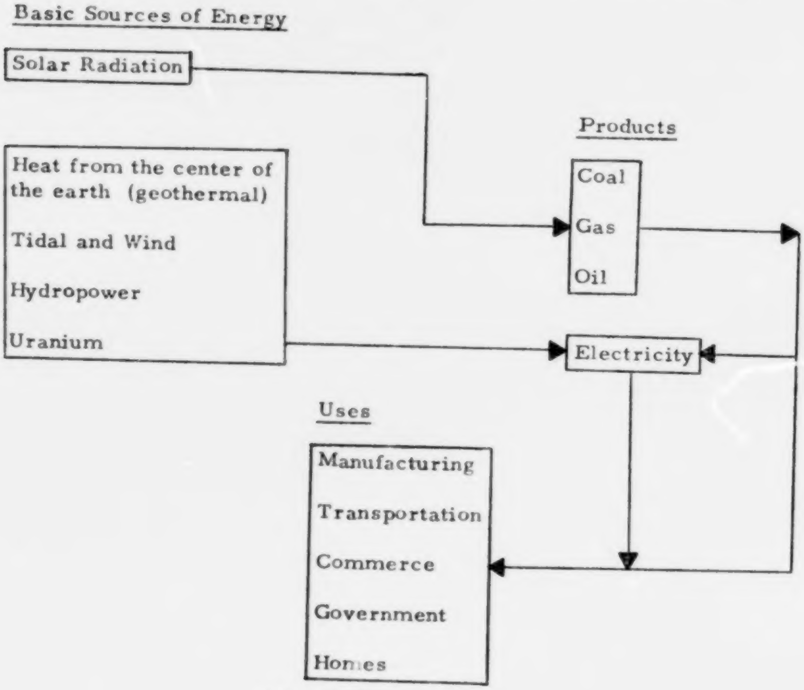

Basic Sources of Energy
Solar Radiation
Heat from the center of the earth (geothermal)
Tidal and Wind
Hydropower
Uranium
Products
Coal
Gas
Oil
Electricity
Uses
Manufacturing
Transportation
Commerce
Government
Homes

CHART 1—*Sources of Energy*

the United States and ever increasing prices as less rich deposits are used. It is of course possible that in the next decade innovations for alleviating shortages may appear such as gassification of coal and breeder reactors. However, the experiences with the long lead times to get nuclear plants into operation cause us to believe that we should not rely on such innovations for the near term.

The several steps in the manufacture of energy are the following:

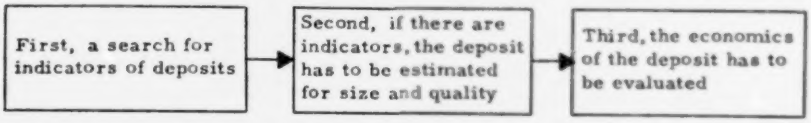
Exploration
Mining
Transportation to energy manufacturing plants
Manufacturing of energy
Distribution

In the remaining part of this section on the industrial decision making, discussions of the types of models used in the above steps will be presented.

Exploration Decisions

Exploration for new reserves of oil, natural gas, uranium and coal is important to the longevity of the companies involved. The use of modeling in exploration is poorly documented because of the confidential nature of the techniques. In the exploration for petroleum there were some early attempts to use game theory to answer questions of moving to drillings in new fields. This yielded poor results and more recently, the methodology of decision theory has been applied. Underground oil surveys make use of other modeling techniques including magnetic, gravimetric, and palynological. The model which combines all these into a comprehensive decision system has not been developed.

Prospecting for uranium ore and for coal generally takes on the following three decision making stages:



First, a search for indicators of deposits

Second, if there are indicators, the deposit has to be estimated for size and quality

Third, the economics of the deposit has to be evaluated

The first stage makes use of all the tools and models of a geological nature. A few attempts have been made in the last two years to use search theory. This generally takes the form of a model to do a rough search to narrow down the likely points, which is then followed up with a more costly search on the selected areas. The use of sampling theory models in the second step has produced forecasts of value in determining where and how close to make test borings. The third step is a market analysis to arrive at expected profit if the ores go on the world market.

Mining Decisions

When a company makes the decision to build a mining operation and exploit a deposit, it frequently makes use of a PERT network analysis to provide a schedule for each of the steps in the construction. It is almost standard practice today to build a simulation of the mining operation, including the operation of the railroad to remove the ore, the operation of a port facility, hourly rates of mining and schedules of the ore ships. The purpose of the simulation is to determine the best schedules for each part of the operation to prevent backlogs and excessive use of storage facilities. For an illustration see [30].

Transportation to Energy Manufacturing Plants

In the petroleum business, the scheduling of crudes from the wells to the refinery plants via tankers and pipelines is a complex problem which must also take into account the ability of the refinery to accept the delivery. Storage costs money and storage capacities are limited. A tanker fleet is composed of ships, each with several compartments for different grades of crudes, and is severely affected by weather delays and difficulty in port entry. A pipeline is a continuous flow of shots of different crudes and is subject to difficulties of predicting arrival times and correct amounts. The requirements of the refineries for different grade crudes is subject to much change. Simulation models have been used to study this area by different companies without significant success and the use of models here is very limited. Some companies have used heuristic models for generating schedules.

Manufacturing and Distribution Decisions

The manufacture of petroleum products is an area where a great deal of successful model building has taken place to support decision making. The earliest use of models was for answering questions on blending. This operation is the mixing of a variety of manufactured components to yield products that meet industrial specifications. For instance, to yield a gasoline of a certain octane and with other specifications, as many as half a dozen different components may be blended. The early blending models were linear programming models using about 40 equations and about 80 variables. The equations were constraints on components, product requirements, specification constraints and an objective function to minimize cost. The blending models generally gave cost savings of about 2% over the manual methods of arriving at the solutions. These models were gradually enlarged to aid in decision making on such problems as desulfurization and catalytic cracking.

Following this phase, more massive models were made of complete refineries to identify the operations and components that would minimize the cost. General opinion seems to be that the cost savings with the complete refinery models was not appreciable. However, the models did provide very valuable information to management on the marginal costs associated with limitations of the plants and finished products.

Petroleum companies have investigated complex models of several refineries. The aid to decision making in this case is the testing of many inter-refinery strategies to see what interchange of components, sources of supply, and allocation of products would minimize the cost of the operation. These have proved most useful for long range planning of facility changes, allocation of product manufacture, and allocation of supply of the different types of crudes. These models have shown considerable cost savings for the companies.

The conversion of energy from coal, oil, gas, and nuclear energy into electricity is characterized by very large scale operations. Electricity utilities are generally joined together in large power pools for the solution of peaking problems and the borrowing of electricity when equipment fails. Utilities require a continuous heavy capital investment. New plants and transmission facilities have very long lead times and are expensive. The optimal configuration of networks and facilities and the efficient use of the equipment are two major areas of industrial decision making where modeling is playing an important role. Utilities normally have the exclusive franchise for the area that they serve, prompting (in recent times) their concern for social and other considerations. For instance, new environmental control measures are causing utilities to consider the social and environmental impact of the location of plants and transmission lines. These restrictions are being modeled along with resultant costs to the utilities and then to the consumer.

Since electricity cannot be stored, the prediction of demand is the starting point for many of the investment decisions that must be made so that the facilities will be available in time to meet the demand. Short term demand estimating models are required to schedule the manning, the start up of generators, and to resolve the decision problems of what to do about peaking. Longer term, ten year estimates, are used to decide on the investments in new plants and transmission lines. Each utility has its own method of prediction, but most of these methods are simple extensions of trend, e.g. the fitting of an exponential curve to the data.

Demand projections are art forms rather than science, because a variety of extensions of trend can be made from past data. Some analysis has been made to assess the risk of shortage if the exponential trend is followed. The large amount of growth, averaging about 7% per year, has so far protected the utilities from overinvestment in plant. Thus the decision making models did not have to be so very accurate. Management's attention has been directed towards reducing the risk of undercapacity to an acceptable level.

Linear programming models have been used to determine the optimal choice in changes of equipment and plant over the long term. These models seek to minimize investment and operating costs using a demand figure produced by the estimating methods. Monte Carlo simulations are used by utilities to assess investment options. The simulations allow modeled equipment to breakdown, allow time to overhaul the equipment, assume demand irregularities, and input of power from the pool. The simulation results allow management to assess a variety of options by simulating their operational characteristics.

A network of plants in a utility complex requires scheduling to find the optimal level of production for each plant to meet the overall system demand. Plants differ in efficiency at different load levels and the loss of about 15% in transmission plays a part. Several modeling techniques including the calculus of variations, dynamic programming and integer programming have played significant roles in providing management with schedules. Linear programming has been useful in modeling the shipment and storage of fuels for the plants. Simulation has been used for modeling the properties of hydroelectric systems for assessing the fluctuations of production resulting from variation in the water sources.

In summary, the use of models to aid industrial management decision making in the production of energy is well developed. The objective in most of the applications has been cost minimization and profit maximization. The questions of what is best for the local community, the State or the nation have not been explicit in these modeling efforts. The growing shortage of energy to meet rising demand and the growing resistance to environmental degradation due to energy production, are bringing new and more complex decision problems to industrial management. This will call for the extension and new applications of models.

III. GOVERNMENT DECISION MAKING AND THE USE OF ENERGY MODELS

The Federal government needs models that will permit energy policy analysis affecting demand, supply and cost. These models, which do not exist in an operational sense today, should be able to provide guidance on matters of what the government should do about import quotas, stimulation of new exploration, conservation of energy and the stimulation of new methods of manufacture of energy. Thus, Federal government decision making in this area is dependent on the estimates of supply, costs and demand. Correct assessment of future energy characteristics is important for planning regulatory, tax, environmental and other policies. It is critical in determining the best allocation of government funds for research and development, as the time from initial commitment to major commercial use may be decades.

In 1970 almost 10% of the useful work in the country was done by electricity. This required 26% of the gross consumption of fossile energy to fuel the plants. The difference is due to the inefficiencies of power generation and transmission. The rate of growth in the use of electricity has been changing from 7% per year in 1961–1965 to 8.6% in 1965–1969 and 9.25% in 1970. The energy consumption for industry, homes and commerce has increased at a fairly steady rate in the last decade, but transportation has had a sharp increase since 1966. The GNP, one of the traditional dependent variables for energy prediction, has fallen off and the ratio of GNP to energy consumption has suddenly started to climb since 1967. These facts illustrate the difficulties of modeling energy consumption by way of simple methods of trend extension.

Battelle Memorial Institute [43] reviewed and compared the significant forecasts of energy for the purpose of assessment of the methods and the predictions. Most of the forecasts studied had only limited information about their methods, and practically none provided quantitative information on the standard errors of estimate or other measures of uncertainty. All of the projections were based on trend analysis or simply judgment. Most of the projections were made from additions of the projections from each energy source such as coal, oil and natural gas. Of the thirteen projections analyzed, the variation was fairly large. For instance, the projections of energy requirements for 1980 varied from 74,542 trillion BTU to 97,000 trillion BTU.

The models analyzed by Battelle are those that influence government policy in the field of energy. One major criticism of these models is that they ignore the relationships among demand, price and supply and they ignore fuel substitutability. Recent model developments are going beyond simple energy trends. For instance, the Federal Power Commission estimates that the demand for electricity will double by 1980 and double again by 1990. These estimates include extrapolations of economic and population growth. The projections will be accurate if the predictions of the economy and population are accurate and the assumed energy consumption relationships continue. More sophisticated models are appearing [2], with alternative scenarios, multivariable relationships with demand and separate estimates for different consumers. We note, of course, that it is possible that the Federal Power Commission estimate could be in error and research on the measure and estimate of such errors is required. For further reading see [1].

Models do exist which analyze the supply and price of different forms of energy [16], [17], and [18]. Models and studies have been made on the demand for energy [21], [22], and [20]. However, these models do not explore in depth the interdependencies of price, supply and demand, nor do they estimate the effects of fuel substitution. For instance, if the price of natural gas should rise, it is to be expected that customers will, to some extent, switch to oil and coal. This ease of substitution is an important omission in the modeling efforts made so far.

Projections by Trend Extension

As an illustration of the basic energy projection model using trend extension and judgment, we cite the National Petroleum Council's study of the nation's energy outlook [13], [14], sponsored by the Department of Interior. The projections were to be made to the end of the century, were to encompass all ranges of possibilities and were to examine areas where Federal policies and programs could contribute to an optimum posture. The study included shale oil, synthetic oil and gas, and nuclear stimulated gas supplies. The description which follows only covers natural gas and conventional oil.

The study was done without econometric and demographic inputs, with no price or production optimizing logic. The computational model computes a schedule of annual oil and natural gas reserves based upon assumed trends in drilling, reserves in fields discovered to data and future changes in oil recovery. From this, the model computes schedules of oil and gas production. The computations also cover annual expenditures required for the exploration, development and production of oil and gas reserves. The model also calculates the annual revenue required for domestic oil and gas production to earn a given rate on investment. Five rates of return are assessed in the model, and for each, five prices of production are calculated.

Fourteen geological regions were analyzed, each having different success rates in exploration, drilling requirements, producing characteristics and costs. The overall model is composed of three submodels, one for oil, one for natural gas, ond one for economic investment calculations. The input to the oil submodel is composed of many factors such as: annual rate of change in exploratory drilling footage, allocation of drilling to each region, years of delay in recovery projects, existing reserves, investment required.

The model consists of accounting type relationships to manipulate the more than twelve input factors for each of the fourteen regions. The output of the oil submodel consists of the following projections: annual exploratory footage drilling, distribution of annual footage to regions, annual footage by well type, oil in place discovered annually, annual reserve additions, annual natural gas discovered, annual oil and gas production, annual reserve additions, annual well abandonments, annual investments, annual producing expenses.

The gas submodel has similar inputs and outputs to calculate future production schedules of non-associated gas and natural gas liquids based on continuation of production from current reserves and trends in drilling and finding of gas.

The economic submodel computes the tables of investments and operating expenses. Using oil and gas investments in net fixed assets at the start of the study, an annual schedule of net fixed assets devoted separately to oil and gas operations is developed. Then the annual revenue required to provide each of five rates of return on the net fixed assets is determined. The required annual revenues from oil operations and gas operations are divided by their production schedules to obtain schedules of revenue per unit of production (price).

The model was developed by many individuals whose experience in the oil and gas operations is extensive. The development of a model using expert judgments on factors and trend extensions is the most simple of all possible approaches. The accounting type equations tie together all the variables using judgment and expert opinion. This illustrates the lack of modeling in the supply area in terms of developed formal functional relationships that allow optimization and insight in the overall process. Usually, when modeling is performed with expert judgment, it is of additional value to know the range of expert opinion on the most significant factors. These ranges can be used to test the sensitivity of the output to the range of likely error in the variables.

A Demand Estimation Model

The Subcommittee on Science, Research, and Development of the Committee on Science and Astronautics of the House of Representatives was concerned about the multiude of energy estimates and their validity and in June 1972, accepted a study [2] which analyzed some of the problems of electricity estimation. The study developed a national electricity demand growth model for testing projection methodology.

It is a premise in this model that the projections of demand for electricity used in the past, which have been successful and are based on population and GNP growth, will not work in the future. Many of the factors influencing the demand for electricity and departing from long established patterns. For instance, the increasing concern for environmental effects is going to force cost increases in the forseeable future marked by the price increase in electricity in 1971. This modeling effort assumes that there are four main factors affecting demand in the future, they are listed in order of decreasing importance: (1) price of electricity, (2) population growth, (3) income, (4) gas prices.

There is wide variation in the published estimates for future cost increases. This model uses two estimates which span the variety of estimates. The first, a National Power Survey estimate of 19% between 1968 and 1990 and the second, a doubling in 30 years. It is assumed that there will be a 13% increase in the price of natural gas from 1970 to 1990. Two estimates of population growth are used. The first, based on a Bureau of Economic Analysis estimate of 1.4% per year, and the second, that the fertility rates will stay about the same resulting in a stable population in 2030. Per capita income growth is assumed at 1.4% per year.

Thus the model uses two alternative projections of population growth and two projections of price, and the other two factors (income and gas prices) are fixed. The projections of demand are made with the equation:

$$Q_{ijt} = A_{ij}(Q_{ijt-1})^{\theta_i}(PE_{ijt})^{\alpha_i}(N_{jt})^{\beta_i} (Y_{ij})^{\gamma_i}(PG_{ijt-1})^{\sigma_i}$$

where i is the consumer class, j is the region, t is the year, Q is the demand for electricity, A is a constant, θ is a time response parameter, PE is the average price of electricity, N is the population, Y is the per capita income, PG is the average price for gas, and α, β, γ and σ are short run elasticities for electricity price, population, income and gas price.

The results of the four combinations of inputs are shown in Table 1 along with a fifth case, assuming no change in electricity prices, and a sixth case with a 24% decline in prices to 1980 and a 12% decline thereafter. The combinations are based on the Bureau of Economic Analysis (BEA), Federal Power Commission (FPC) the estimate from the National Power Survey of the FPC and zero population growth to 2035 (ZPG 2035).

Some conclusions that can be drawn from this output are the following:

1. At 1975 there are small differences. Supply problems by 1975 will not be eased by adjustment of prices.
2. The population assumption is relatively unimportant for the next 20 to 30 years.
3. The generally accepted projections and estimates of the influence of causal factors (price, population) are incompatible.
4. Previous projections of demand growth from FPC, National Petroleum Council, and others were too high.

This research was prepared for a committee in the House of Representatives, and was aimed at an investigation of the multiude of estimates of the energy crisis. The predictive model that was designed, an iterative year by year calculation process, allows the input of many assumptions about causal variables. It is assumed that demand for electricity can be projected without regard to the demand pattern that will develop for the other fuels.

A Dynamic Demand Model With Interfuel Competition

We next describe a model that combines the supply, demand and price in a

Table 1 Electricity Demand Forecasts

Population Assumption	Price Assumption	Electricity Demand *			
		1975	1980	1990	2000
Bureau of Economic Analysis	Federal Power Commission	1.98	2.38	3.01	3.45
Bureau of Economic Analysis	double by 2000	1.88	2.07	2.11	2.01
Zero Population Growth 2035	Federal Power Commission	1.98	2.37	2.95	3.29
Zero Population Growth 2035	double by 2000	1.88	2.05	2.07	1.91
Bureau of Economic Analysis	constant	2.02	2.54	3.56	4.56
Bureau of Economic Analysis	decline	2.14	3.05	5.66	9.89

* Trillion Kilowatt Hours

medium to long range dynamic model of interfuel competition for the United States [7]. This model, developed by M. L. Baughman, has not been used in a decision making framework as yet, but agencies of the government are reviewing its applicability. It is included here because it represents a major advance in energy modeling and is likely to be used extensively in the government. The results of the model will be of direct use in policy analysis.

The model contains the interactions between supply and demand within the market clearing process. The market clearing process yields the price and quantity of the different fuels. For the market to clear, supply and demand must be equal. The overall model is sketched in Figure 1.

The demand by sectors and the resource characterization are the exogenous inputs. Sector demands are time series for transportation, space conditioning, industrial processing, etc. Some part of the total sector demands will be going to the market place to buy energy from the market sensitive demand part of the model. These inputs are functions of the rates of growth of demand and rates of turnover in consumers equipment. The total of consumers who continue to use the same fuels is called the base demand.

The market clearing process, based on price, matches supply to meet market sensitive demand in the classic economic supply-demand equilibrium. This is related to supply schedules, demand schedules and a theory for the market clearing process.

The dash line in the schematic indicates the boundary of the system that was modeled. It is assumed that the variables inside the model do not affect the variables outside the model. The assumption that price inside the model does not affect demand by sectors implies that demands are elastic. Additional assumptions are that if all prices change proportionately, the level of demand will not change in that there

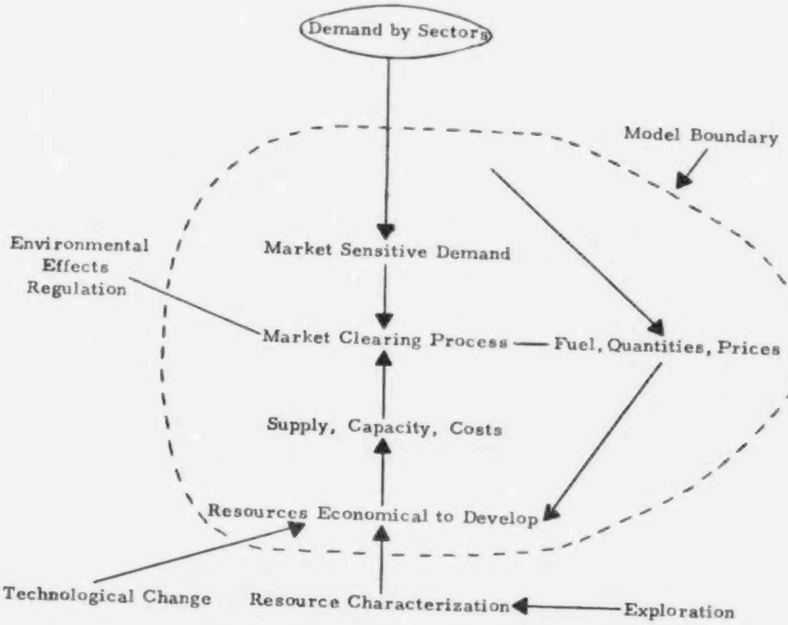

Demand by Sectors
Model Boundary
Environmental Effects Regulation
Market Sensitive Demand
Market Clearing Process
Fuel, Quantities, Prices
Supply, Capacity, Costs
Resources Economical to Develop
Technological Change
Resource Characterization
Exploration

FIGURE 1—*Dynamic Demand Model Interactions*

will be no substitution, and that levels of demand are dependent on variables outside the model such as GNP, population and other demographics.

In addition, exploration activities and the effects on reserves are not dependent on the model variables. The relationships between investment in exploration and the resulting returns on investment are not well understood. Another assumption is that there is perfect competition in that prices of supply are equal to cost. It is hypothecated that this is reasonable in spite of regulations and the fact that imperfect competition does exist. It is also assumed that import and export quotas are exogenous variables. Electricity which consumes fuels and competes with them on the market is treated as both a supply and a demand with a price dependent on fuel price and process cost.

The level of aggregation is at the coal, oil, natural gas and electricity levels of supply. Thus prices are really price indices over the subcomponents of the supply aggregates. This creates a loss of sensitivity to the changes in demand for subcomponents and resultant changes in production. It also assumes that there is an adequate shipping system for all products, and that costs can be treated afterwards. Geographical considerations are not included in the model.

One output of the model is the market shares and levels of consumption for the fuels and electricity. This output might aid in developing policies on pollution, as well as energy. Another outut, the supply and price of energy, might be valuable to the government in terms of predicting economic growth. The levels of investment in the fuel suppliers, as a model output, might be helpful for predicting energy costs and the availability of investment for other fields. To illustrate the level of mathematical statements used in the model, the following are the cost functions for the cost of petroleum.

Given an oil reservoir developed to capacity q_o, the output would appear as the solid line below, in Figure 2:

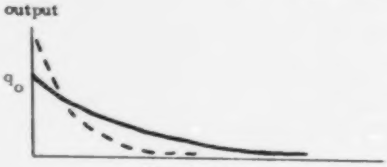


FIGURE 2

The dotted line shows increased capacity. This can be approximated by an exponential $q(t) = q_o e^{-Dt}$ where D is the rate of decline in output.

Using this formula, the total capacity of the reservoir can be estimated. The government policy decisions on investment in development can then be assumed. These include:

- money for speedier recovery of the oil
- investment for a more complete recovery of the site

These options can be tested by formulas that yield the marginal development cost, which is the cost of the next barrel resulting from investment in more capacity, i.e.

$$MDC = \frac{b(q_o + r_o R_o)^2}{r R_o}$$

where

MDC = the incremental cost of the next barrel resulting from investment in more capacity,
r_o = the discount rate,
R_o = the total resources in the reservoir,
q_o = the initial output of the reservoir, and b and r are constants.

The MDC formula is really a submodel in the overall sets of equations. It provides the marginal development cost for each of the overall tests of the model. The overall model is an equilibrium seeking, dynamic system that searches for the equilibrium of supply and demand.

The model has been validated by using it to predict the behavior of prices, demand and supply of the several fuels over a historic 50 year period. The match to the historical data was considered good.

This model, as illustrated by the example of finding the marginal development cost of petroleum, does not need a great deal of data input since most of the activities of price, demand and supply are represented by formula. The many assumptions and conditions noted previously suggest that this model requires further development and validation. However, it is felt that its use by the Federal government in its present form would be helpful for making applicable decisions as there is a definite lack of valid other approaches.

A Pollution Policy Model

The previous model is significant in energy model building for decision making purposes because of its comprehensive characteristics. As noted, one of its outputs is the market shares of the several fuels and the levels of consumption. This can be used for pollution analysis and testing government policies on pollution. Another model, which we describe next, is the Energy Quality Model [8], being developed for the Environmental Protection Agency for testing pollution policy.

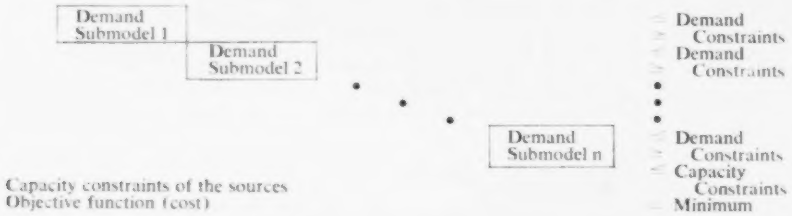
This effort was aimed at determining the effects of sulfur limitations on the allocations of fuels to regions of the country. A mathematical model was developed for predicting fossil fuel supply, demand, cost and distribution for the air-quality control regions subject to sulfur limitations. The approach was to link together three submodels on demand, transportation and bidding, then solve for fuel prices and allocations to regions over the entire United States.

The model is formulated as a large linear programming problem. Ten types of fuel covering the types of oil, coal and gas have a total of 300 sources of supply. The model is to answer two questions:

(1) How do the 300 fuel sources get allocated to the air quality regions so as to minimize cost and still meet the specified air quality constraints?

(2) What is the price for each fuel type in each of the regions?

The linear programming problem can be outlined by the following diagram.



The model requires as input the following:

- Selling prices for the fuels in each of the regions
- Values describing the physical characteristics of the power plants, Btu requirements, and sulfur emission limits for each region
- Demand for coal, gas and oil by the non-power plant users.

The output of the model is the following:

- A fuel price for each fuel used in each region
- The amount of each fuel used in each region. This model does not include the use of fuel for transportation in each region, but does include commercial, home and industrial usage
- The amount of each fuel shipped from each supply district to each use region.

This is intended to be a predictive model in that it is to be used for estimating the effects of sulfur policies several years in the future.

Demand for each of the air quality regions is assumed to be fixed for each of the fuel types. Fuel cost is estimated by adding production costs and profit. The output will also include the cost of transportation. Power plants are allowed fuel options including fuel switching from type and grade, and the installation of gas cleaning equipment. These costs are added to the fuel costs.

As an illustration of the type experimental runs, a predictive run for 1975 was desired. The AQCR's in this case were aggregated to match states so that the state implementation plans could be studied. Some AQCR's were in two states, requiring fragmentation. In addition, 12 critical metropolitan regions were included. The final aggregation included 46 states, 10 metropolitan regions and 5 fragmented metropolitan regions. Fuel supply districts were aggregated to 112 regions. This illustrates the flexible nature of the model in that it can be aggregate or disaggregate as required by the study.

The Energy Quality Model is still in the developmental and testing phases and, as far as can be determined, it has not yet been used for decision making or policy analysis. The model should be of value in pollution policy analysis by testing the effects of pollution constraints on the national or regional distribution of fuels and the resultant prices to the consumers. The model is significant because it points the way for modeling of interactions of environmental quality policies and the effects on society in terms of the availability and price of fuels.

A Model for Rate Making Proceedings

The procedural steps in Federal regulation are defined by law, but the processes for each of the steps are not clearly spelled out. The use of econometric models for future decision making is a real possibility. In 1961, the introduction of an econometric model on the Permian Basin Area Rate Case was the first use of the results of a model in a rate case before the Federal Power Commission [10]. The model projected the impact that various ceiling prices would have on the demand and supply of natural gas.

The model as conceived by the economics staff of the Federal Power Commission, used as inputs an assortment of hypothetical gas prices which resulted in estimates of future exploratory effort, additions to gas reserves, gas production, gas consumption and other factors affecting the natural gas industry. The model was developed in three separate parts: exploration activities, residential and commercial consumption, and industrial consumption.

The basic analytical approach was regression analysis. Each of these parts was represented as a dependent variable in the regression approach. For example, to create a dependent variable for exploration activity, the Office of Economics selected the number of exploratory wells. The associated independent variables included the expected future price, the production of oil and gas, the cost of money, and exploration success ratio, an index of the probable depths of new reservoirs and finally, a grouping of many factors in one index. These variables were identified for the period from 1940 to 1961 and were inputs to the regression analysis. The residential-commercial and the industrial submodels were developed in a similar way.

The basic finding of the model implementation was that when there is a gas price increase there is a tendency to lowering of the quantity of gas sold. The supply of gas is essentially the underground inventory of proven gas reserves. Inventory levels are referred to as life indexes in which proven gas reserves are divided by annual gas consumption to determine how many years' gas supply is available at any given time. With reduced consumption, after a price increase, the life index value would rise, meaning that more gas would be available without new reserves. The conclusion is that price increases do not stimulate new exploration.

The model was severely criticised by industry experts and others. Oil and gas producers felt that the model was inaccurate, full of subjective manipulation and not a useful tool. The hearing examiner felt that the model was not relevant or material to the problem. The subjectivity, the problems of multicollinearity, bias, problems of autocorrelation and aggregation were all spelled out as rendering the modeling inadequate as a predictive tool.

A Model for Federal Procurement of Fuels

One of the decision making areas for the Federal government is to determine the sources for the purchase of fuels for use by the agencies of government. The Defense Fuel Supply Center (DFSC) has been purchasing fuels since the mid-sixties using a linear-programming model for the selection of bids from suppliers for delivery all over the country [14]. The procurement is for the entire defense establishment. It is the world's largest purchaser of fuels, and the U. S. petroleum industry's biggest customer. This case history involves the jet fuel procurement program which accounts for more than half of the one billion dollars DFSC spends each year.

DFSC procures more than two billion gallons of jet fuel for the armed forces every six months. It must be shipped to fill the requirements of over 300 air bases, air fields, and air stations on the continental United States. Every six months DFSC must consider about 100 bids from petroleum companies throughout the United States and South America, and award contracts to the companies so that the total delivered cost is the least possible.

Although the total volume of fuel offered by all bidders exceeds the total requirements, it is never possible to award contracts such that the requirements of each base can be completely supplied from the least expensive source. Import quotas and limited offerings by each bidder preclude this. Simple procedures of awarding contracts to the cheapest sources, then to the next cheapest sources for unfilled requirements, and so on, do not provide the least-cost solution.

As an illustration, suppose there are four bidders, labeled A, B, C and D and three supply depots or destinations labeled 1, 2, and 3. The four bidders offer the following quantities of the same product:

Bidder A offers a maximum of 800 units

Bidder B offers a maximum of 1200 units

Bidder C offers a maximum of 1100 units

Bidder D offers a maximum of 3000 units

Total is 6100 units.

The three supply depots require the following quantities of the product:

Supply depot 1 requires 1000 units
Supply depot 2 requires 1500 units
Supply depot 3 requires 2000 units
Total is 4500 units.

The bid price per unit of the product by each bidder, fob destination, is indicated in the following table:

	Supply depots			
	1	2	3	
Bidder A	52	30	45	**Bid prices**
Bidder B	48	32	50	
Bidder C	60	40	55	
Bidder D	52	41	60	

According to this table, the price per unit bid by A and shipped to depot 1 is 52 cents. The cheapest source for depot 1 is bidder B at 48 cents, for depot 2, it is bidder A at 30 cents and for depot 3 it is also bidder A. Bidders A and B do not offer enough product to satisfy the needs of the three depots. Other bidders must be considered. If the reader attempted to award bids even in this simple case, he would find that the least-cost set of awards totaling $2,051 is difficult to arrive at.

In addition to complexities of amount of product, a bidder may indicate a minimum quantity, indicate a choice of shipping points, tie his offer to another bidder, quote different prices for different amounts and many other complexities. The awards must be made in conformance to the Small Business Act, import quotas, and the criteria may be fob origin or destination.

The problem is treated as a standard linear programming transportation problem, i.e.

$$\text{Minimize} \quad \sum_i \sum_j C_{ij} X_{ij} \quad \text{(total cost)},$$

$$\text{subject to} \quad \sum_j X_{ij} = r_i \quad \text{for all } i$$

$$\sum_i X_{ij} \leq b_j \quad \text{for all } j$$

$$X_{ij} \geq 0 \quad \text{for all } i, j$$

i denotes the installation
j denotes the bidder
r_i = the requirement by installation

b_j = the maximum offered by bidder j
X_{ij} = the quantity shipped to i from bidder j
C_{ij} = the delivered cost per gallon to i from j

The above formulation does not handle certain possible bidding strategies, e.g., increasing for additional lots from the same bidder, but mathematical and computational procedures can account for these complexities. The system using the transportation method of arriving at the least cost had reduced total cost by more than 2 percent, reduced the bid evaluation time to a 1 hour run on the computer, allowed quick reaction to changes, eliminated computational errors, and the model can be used for a multiplicity of planning studies.

A Model to Locate Needed Research

A significant function of the Federal government is the awarding of grants and contracts to solve technical problems in the energy field. The purpose is to produce innovations to increase supply or to increase efficiency of the operations. Models of systems can be very helpful in identifying particular aspects of operations that would benefit from research. The models can also be used to determine the benefits from changes in technology in terms of greater production, reduced pollution, lower cost and other features. The following model is one of a series designed for this purpose [6]. It is a techno-economic model of an electricity generating plant, using coal, gas and oil. For a variety of such studies, including locational studies, see [23], [24], and [26].

At present, fossil fueled steam electric plants produce 80% of the electrical energy in the United States. The fruits of R&D in cost reduction in these plants would have a substantial effect on the economy. This model is intended to predict cost reductions based on hypothesized advancements.

Since the model is to assess technological advancement, it must contain technical detail of sufficient depth. For example, to assess the cost reduction resulting from improved heat transfer coefficients in the condenser, the model must be able to include detailed computation of the size and related cost of the condenser for various heat transfer coefficients.

The first step in the modeling effort is to make a schematic diagram of the typical plant, Figure 3. For each of the elements of the schematic, cost equations are developed related to design characteristics. To demonstrate the methodology, one component for water treatment, a wet tower with natural convection will be made into a cost equation.

The method to develop the equation is to use known engineering relationships of performance to produce the following:

$$\text{Cost} = (0.015)(X)(Y)(0.8),$$

where X is the heat removed from the water by the tower, Y is a performance coefficient as a function of the design characteristics. Using regression methods, the validity of the cost equation was checked against actual results of specific power plants.

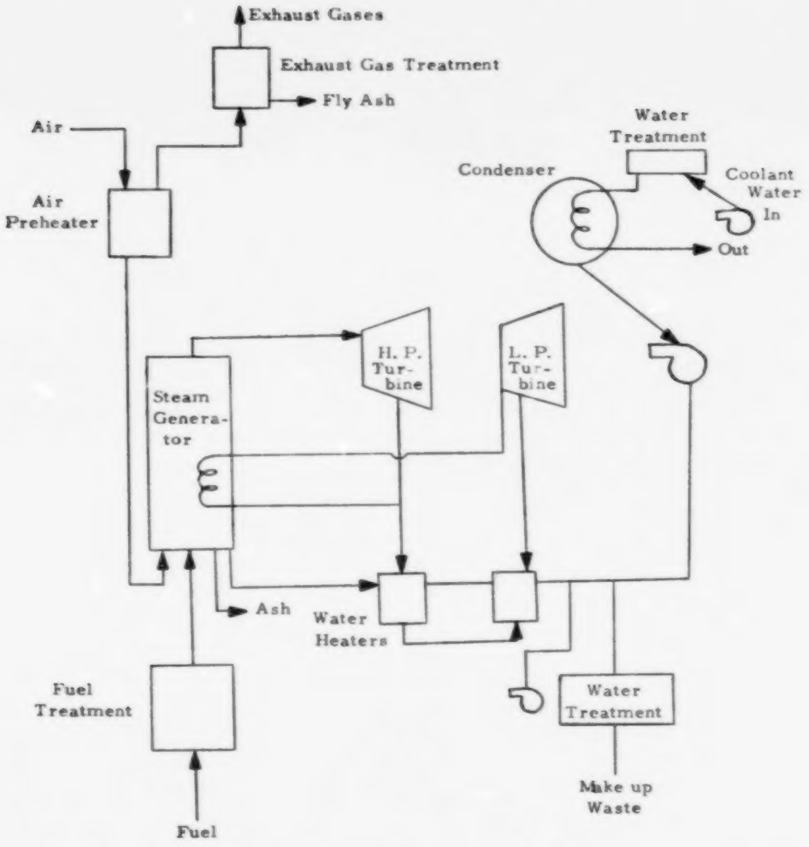

Exhaust Gases
Exhaust Gas Treatment
Fly Ash
Air
Air Preheater
Water Treatment
Condenser
Coolant Water In
Out
H. P. Turbine
L. P. Turbine
Steam Generator
Ash
Water Heaters
Fuel Treatment
Water Treatment
Make up Waste
Fuel

FIGURE 3—*Steam Electric Plant Schematic*

The outputs of the model are based on postulates for engineering advancements (changes). Table 2 illustrates the typical output.

The usefulness of models of this type, aimed at identifying significant research projects for engineering research, is limited by the simplifications which must be made. As far as we can determine this model has not been used to evaluate alternative decisions.

A Model of Oil Supply and Distribution

As has been illustrated above, the Federal government has depended on simple trend analysis methods for determining supply, price and demand. Interfuel competition has not been analyzed in the models used thus far. Of late, the Baughman model and the EPA Energy Quality Model have shown promise of increased quality of modeling. The use of the recently developed models described in this paper in decision making appears to be small or non-existent. However, agencies such as the National Science Foundation, Environmental Protection Agency, Council on Environmental Quality and the Federal Power Commission are funding efforts to develop modeling approaches that will be of direct use in answering specific decision problems.

Table 2. Engineering Advances vs. Cost Savings

Advancement	Result	Coal	Oil	Gas
		Savings-mills/KWH		
Reduce size and cost of cooling tower by 20%	Lower construction costs	.5375	.5086	.5037
Increase boiler outlet temperature from 1000° to 1400°	Reduction in fuel consumption	.0604	.0604	.0604
Eiliminate fouling in condenser	Reduction in condenser size	.117	.117	.117

In contrast, the Canadian National Energy Board has an Operations Research Branch which has developed several models that are contributing to that nation's policy decision domain, see [15], [46], [47], and [48].

A model conceived for the National Energy Board was used in the assessment of the effects of the Arctic oil findings on the Canadian oil industry. Although the decisions that were made based on use of the model have not yet been divulged, and even though it is rather complex, we feel that the model is an important one and we next discuss its structure.

The problem called for a global solution at a macro-economic level and in addition, answers at a micro-economic level concerning linking of supply and demand distribution points by pipelines to include pipeline capacity, cost, effect on field prices, and investments and producing capacities in the various regions. It was assumed that profit was the prime motive of competition between producing centers.

Major assumptions were as follows:

- Price changes are triggered by the leader in the industry who is the major operator in the producing region.
- The Arctic region will be able to compete with other continental regions on its own terms in order to maximize its profit.
- The base price is set by the price leader at the most competitive producing region, determines the maximum value or opportunity cost of crude at all other points in the distribution system.
- The cost to find, develop, and produce oil in a region has little bearing on the posted field price of oil in this region.
- Price is a determining factor in the decision to invest in oil producing capacity.
- The oil distribution system tends to minimize the cost of fuel delivered at the demand centers.

The data base is quite extensive and consists of five categories of information.

1. By 1° latitude and longitude differences, detail of terrain and province boundaries.
2. Regional cost of pipe, mean winter temperature, and marginal cost relationship characterizing the economics of exploration and production.
3. Sixty supply and demand nodes are identified by latitude, longitude, and characterized by additional data. Supply nodes are characterized by field price ber barrel, a list of regions for which the node acts as a collection point, and the production of the regions. Demand nodes are characterized by demand levels and growth factors.
4. Pipeline route data are piece-wise linear curves, a chain of straight line elements. Each link on the chain is characterized by the latitude and longitude of the end points, the capacity of the link, the direction of flow, cost of transportation and the pipeline diameter.
5. Parameters which include the discount rate, cost of servicing

pipelines, oil import quotas, price of steel and oil and similar factors.

The demand for oil and the field price is related to a level of exploration and development by a production function. The allocation of crude from supply to market is constrained by the producing capacities and by oil import restrictions. These capacities are a function of recoverable reserves which in turn are a function of investments in exploration and development. These investments depend on the economics of exploration and development, price of crude and the cost to find, develop and produce oil. While the price of crude is a result of competition, the cost to find and produce oil depends on the available geological information.

The model requires an estimation method for the amount of investment in exploratory and development operations. This is done by first deducing the relationships among probability of finds, estimated volume of basins, depleted and discovered basins and hydrocarbon reserves. This resulted in an equation of investment between the current known year of data and the year n. This was then expanded by integration to a linear equation which was approximated by regression analysis.

A capacity expansion function was devised to estimate regional investments in exploration and development for oil and gas, based on field prices and demands for oil, as determined by the competitive position of the respective producing regions in the supply distribution system. A choice had to be made between the traditional econometric regression model and a deduced behavioral model. The poor results of regression and the lack of data indicated that a deduced model was the best choice. The model was validated against 100 aggregate annual investments for 40 producing regions and was adjudged satisfactory for the purpose of the overall model.

Finally, a flow allocation model was developed. At every yearly time step, this model solves for the minimum cost allocation of oil in the supply and distribution system and determines which existing pipelines should be expanded and which new lines should be built. The distribution network is modeled as a set of capacitated arcs. This is transformed into a network flow graph amenable to mathematical flow optimization procedures. Figure 4 shows a simplified network. Dummy demand and supply nodes are included to absorb any slack in the system.

All demand centers are connected to a dummy demand center with imaginary pipelines having minimum and maximum capacities equal to the respective demands at each center. These links incur no costs of transportation. All supply centers are connected to a dummy supply center with imaginary pipelines with maximum capacities equal to the producing capacities of the supply points and with minimal capacities equal to zero. The unit transportation costs through these imaginary pipelines are the field prices of oil at the respective points of supply.

Other imaginary nodes and links were necessary to describe certain conditions of the system. For example, the tariff imposed on Canadian oil exported to the United States is represented by introducing two imaginary points at each side of the border along pipelines crossing the border. The trans-border link between these two nodes is assigned a zero length, an unlimited capacity

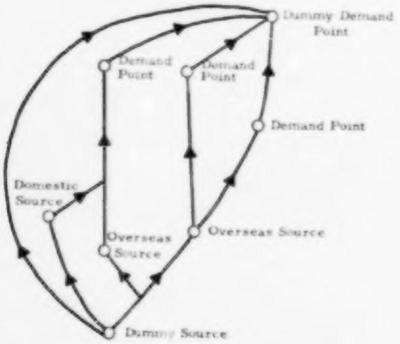


Figure 4—*A Simplified Version of a Flow Network*

and a positive import tariff for southbound pipelines.

The system is then represented by a graph N composed of arcs (i, j) each characterized by: a maximum capacity Q_{ij} going from i to j, a minimum capacity M_{ij} going from i to j, a unit cost of transportation C_{ij}. The optimization problem is to find a way to solve for a flow amount q_{ij} on each arc that minimizes the total cost satisfying the minimum capacity constraints and subject to the maximum capacity constraints. This can be stated mathematically as a special linear-programming problem as solved using computational procedures which allow for fast and efficient solution of network problems.

An iterative procedure for solution was adopted because the costs C_{ij} in reality are dependent on the q_{ij}. This procedure consists of correcting costs in a second run through the computation based on the q_{ij} values from the previous run.

This model is significant because it points the way for the United States to model the supply points and the demand points in such a way as to allow testing of alternative arrangements of supply and distribution facilities. In the years to come, when shortages of oil become severe and prices increase, the need for a more efficient national distribution system will assume increased importance.

A State Government Demand Model

State governments have not used modeling to any great extent in their planning and decision making. One noteworthy exception is the State of California, which has developed several energy planning models. The model described below was aimed at providing a basis for a twenty year power plant citing plan in terms of projected demand [9].

One possible method of building a State model that was considered was to simply add together the projections of all the utilities in the State. It was decided that this would not be practical because each utility has variations in their methods, but all methods amounted to simple trend projections. Since the growth rate is generally exponential, it was felt that this cannot be realistic for future projections. Modifications of this rate would have to occur.

The model approach used was to view the State as a whole and develop a forecasting model with social, demographic, technical, geographical and other factors. Five cases were developed for possible future conditions, all reasonable projections of economic and demographic growth and with some anticipated increases in the price of electricity. The likelihood of any one of the cases actually occurring cannot be evaluated; however, there was a base-line case which is classed as a surpirse-free scenario. The base-line case is bracketed by two cases, one with lower estimates and the other with higher estimates. The remaining two cases of the five accounted for the possible effects of increasing energy prices. The five cases are as follows:

(1) Base case
(2) Base case with price increases
(3) Low-growth case with price increases
(4) Low-growth case with constant prices
(5) High-growth case with constant prices

The model outputs and therefore the conclusions are, of course, sensitive to the structure and relationships of the model, as well as the assumptions in the scenarios. For instance, a few of the assumptions in the base case with price increases are as follows: in the early 1970's pollution restrictions have stopped construction of all fossile fuel plants in the south region; nuclear plants have been obstructed by conservationists and safety considerations; only limited development of geothermal capacity takes place; and by the 1980's brownouts and blackouts are frequent and may last an entire day.

The model represents all of California and is not a summation of different utility marketing areas. It does account for different meteorological conditions and population densities. Price increases in the model effects demand by causing shifts to other energy sources and decreased consumption.

The model consists of a set of six main submodels for the following components: residential, industrial, commercial, transportation, government, agricultural.

The residential submodel is composed of three main factors: number of households, devices in the homes, consumption of electricity by devices, which are functions of demographic change, factors affecting use of appliances, climatological factors, relative prices of energy sources, dwelling unit characteristics, utility promotional activity.

The industrial submodel is composed of 21 industrial group models. Consumption is measured as a function of the output of the group and the electricity required for that output. Assumptions are made about the changing technology, e.g. canning will be supplemented by freezing processes.

The commercial submodel involves five elements of trade, finance, insurance and real estate, communication, electricity, gas, sanitary services, construction. The method of estimating the consumption for the commercial submodel is based on regression with the gross State product (GSP). The future is estimated by simple trend projections for GSP.

Transportation is such a modest consumer, that the submodel reduced to an assumed constant. The agriculture requirements for electricity are related to the electricity used for irrigation which is fairly stable. A constant factor was also used for this component.

Government uses include street lighting and office buildings. The method of estimating consumption is by a relationship of the government contribution to the GSP.

This energy prediction model, as has been noted, is a set of estimating equations where the variables are input factors arrived at from source material on projections of the economy, technology, demographic characteristics and other factors. It is a deterministic model, however the use of the five scenarios allows analysis of possible errors of estimate. The output of the model of course is the demand figure for the State. Model results, running the five cases have led to the following conclusions:

1. Demand could be much lower than would be projected by simple extrapolation of trend such as each utility is doing.
2. Price increases will mainly affect the commercial and industrial sectors.
3. Demand is very sensitive to economic and demographic variations.

IV. RECENT DEVELOPMENTS AND NEW DIRECTIONS

New modeling research efforts are underway in the United States stimulated by the growing concern with the problems of energy. Resources for the Future (RFF), Inc. conducted a seminar in January 1973 where many of the new efforts were presented. The presentations are briefly outlined here to indicate the various directions that current research is taking. It should be noted that the work outlined below does not include all research on modeling energy problems; they do point up many of the new approaches that are being investigated.

Use of Input-Output and Econometric Techniques for Energy System Modeling

The main drive of models of this type is the relating of energy consumption and/or production to comprehensive economic models. This approach, using many economic sectors, allows more than the usual detail modeling of residential, commercial, industrial, transportation and electric utility sectors. The availability of data delimits the extent to which the model can be detailed. Different investigators use various devices to disaggregate historical energy statistics. The way in which energy consumption is related to the economic model for forecasting is one of the features of interest. This can be simply a constant relationship between energy consumption and economic activity. Usually data indicates a trend and more complicated relationships must be used. At the other extreme, full interaction between the energy model and the economic model (in the equilibrium concept) with energy prices affecting energy production and consumption as well as production and consumption of other goods and services. Several models were presented at the RFF seminar. In a paper by William A. Reardon [50], the disaggregation is carried to much greater detail than previously in the construction of input/output tables. Reardon's tables covered three years of time and 35 economic sectors. The trends found for the input/output coefficients indicates trends for energy consumption per unit of output in the 35 individual sectors.

In an Inter-Industry Forecasting model, under development at the University of Maryland [51], there is a division of the economy into 185 industries. The model is used to estimate the sales of each industry to each of the other industries, 112 types of capital investment, consumers, government and export. More than 1000 regression equations were employed in forecasting consumer demand, investment, export and import behavior, labor productivity and changes in materials used in the 185 industries. Forecasts are generated year by year for a ten-year span. In order to forecast demand for petroleum, the model relates each major petroleum product to each industry. In a paper by Verleger [52], there is an attempt to more fully couple the energy and economic models. He makes use of partial and general equilibrium models in which energy and nonenergy demands of the economy are put in a simultaneous equation framework where prices are determined so that all markets clear at every moment of time.

Use of Linear Programming in Energy Models

Linear programming provides a basis for another class of energy models generally making use of the classic transportation problem. Since this is a class of optimizing models, they are not particularly useful for projection. Linear Programming models are attractive to research on planning and policy analysis. Linear programming features the ability to use constraints such as production capacity limitations. Thus, the outputs of the model can be used as standards to measure the efficiency of a given or proposed system. Hoffman [53] offers a model for planning and technology assessment. The model encompasses the entire energy system and represents full inter-fuel substitutability. The analytic approach is to use n alternative supply categories, yielding $n \times m$ possible supply-demand combinations. The current research has n at 13 and m equals 15. The objective function is the minimization of cost bounded by the demand and supply constraints. Another linear programming module by Deonigi and Engle [54] is concerned with the generation of amounts of electric power which is specified exogenously. The model equations are formulated as electric utility decision variables, yielding a model 4,000 × 9,000 in size. Griffin [55] introduces a model applied to the petroleum refining industry using a linear programming description of the production process. The demand estimates are based on conventional econometric approaches using stocks of petroleum and consuming equipment. Once the demands are obtained for products, the requirements for the major products become constraints. The objective function seeks to minimize the cost of production using demand, quality specifications and capacities as constraint sets.

Econometric Models of Individual Sectors

Several papers given at the RFF seminar are interesting because of their simplicity. A paper by Erickson, Spann, and Ciliano [56] uses five consumption sectors with consideration of the role of each fuel in each sector. The model is a regression approach relating demand to a series of variables covering new uses, prices, density of urbanization, income and temperature. In another research paper, by Spann and Erickson [57], there is a focus on long-run oil and gas exploration using econometric techniques. The model seeks to estimate the response of supply to econometric factors such as prices, interest rates, and time. The model contains cross-elasticities of supply between oil and gas and cross-elasticities with demand. Gas and oil are produced and discovered by the same companies in actuality, thus they can be considered as joint products. The supply and demand interdependence is taken into account by using a simultaneous equation framework.

The models outlined above represent the classes of approaches that are being experimented on in current research. Whether input/output, dynamic equilibrium, or linear programming, the trend seems to be models of considerable complexity, requiring extensive data. Interestingly enough, modeling in transportation when challenged to produce better models went to models of great complexity and high data cost. During the last decade, the transportation models have become more pragmatic and simpler approaches are again being used. It would appear that the funding of these research efforts in the energy field should nurture both complicated and simple models to insure a usefulness of the research results.

However, the usefulness of modeling to examine some of the newly arrived problems of energy is dependent on the development of new comprehensive and interactive techniques. For instance, the impacts of new energy technology on the economy and air quality was treated with an input-output model by Just [74]. This model was used to examine three new technologies that might be significant in 1985:

(1) High BTU coal gasification;

(2) Low BTU coal gasification;

(3) Gas turbine topping cycle (combined gas and steam cycle).

The major results show the sensitivity of total capital investment to changes in the energy use growth rate and to the adoption of new energy technology. They also show that very small changes in the overall growth rate of personal consumption or government spending can restrain total investment to within its historical limits as a percent of GNP. Thus the United States can sustain the high investment costs created by rapid energy demand growth by reducing the growth rate of consumption and government spending by less than 0.1% per year through 1985.

The core of the model contains the actual and projected input-output structures describing the 1963, 1970, and 1980 economics. The noneconomic quantities are referred to as accessory variables and are summarized in the bottom half of the Figure 5. These are model outputs and are assumed to be proportional to the total output of each sector.

The boxes in the upper half of Figure 5 show the various means for interacting with the model, in terms of the alternative future being investigated, including changes in technologies and composition (or size) of GNP.

Alternative futures are constructed by developing a final demand vector to represent the conditions of the scenario and modifying the technology and capital coefficients to include the amount and kind of new technology that is specified. Once these changes are made, the total outputs and investment requirements can be calculated. The values of the accessory variables are then obtained by simple multiplication.

In making projections of final demands, it is necessary to calculate the investments required to support the

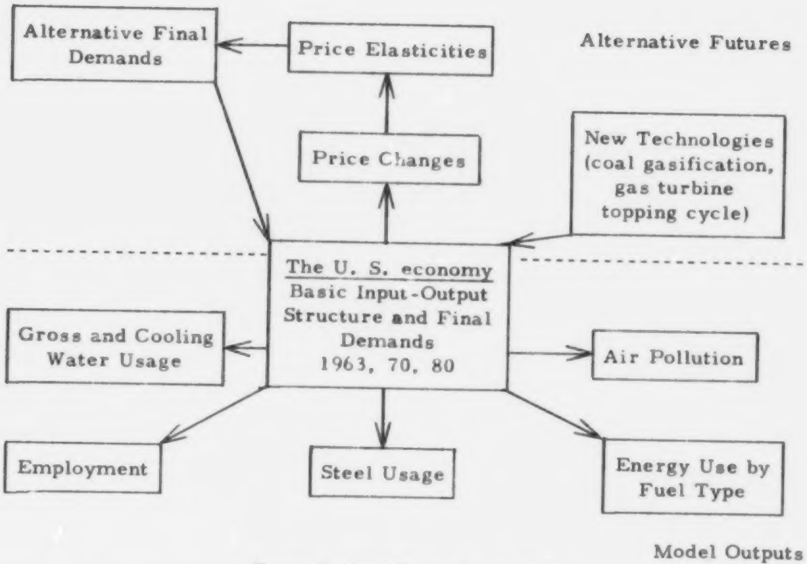

Alternative Final Demands
Price Elasticities
Alternative Futures
Price Changes
New Technologies (coal gasification, gas turbine topping cycle)
The U. S. economy
Basic Input-Output Structure and Final Demands
1963, 70, 80
Gross and Cooling Water Usage
Air Pollution
Employment
Steel Usage
Energy Use by Fuel Type
Model Outputs

FIGURE 5—Input-Output Energy Model

level of final demand, which depends on investment level.

A simple two-period model can be used to illustrate this. Assume that:

(1) The same technological coefficient matrix A applies to both periods.

(2) Total final demand Y consists of final demand purchases by households and government (Y^F) and capital investment purchase by all sectors of the economy (Y^I), thus,

$$Y = Y^F + Y^I$$

(3) The capital matrix C is defined as $C = (c_{ij})$ where c_{ij} is the marginal capital purchase from sector i by sector j, required to expand the capacity of sector j by one dollar of output. Thus, if X_o is the total output in period T_o and X_1 is the total output in period T_1, the new investment required is $C(X_1 - X_o)$.

The objective is to find for period t_1 the total output (X_1) and total final demand (Y_1), given the total output in period t_o, (X_o) and the noninvestment final demand in period t_1, (Y_1^F).

The basic equations for the model are:

$$X_1 = (I - A)^{-1} Y_1 = (I - A)^{-1} (Y_1^F + Y_1^I)$$

and

$$Y_1^I = C(X_1 - X_o)$$

These can be solved for total output (X_1) and total final demand (Y_1):

$$X_1 = (I - A - C)^{-1} (Y_1^F - CX_o)$$
$$Y_1^F = X_1^F + C(X_1 - X_o).$$

These can be used to investigate the effect on investment Y^I and the total output X of changes in the growth rates of components of Y^F.

Projections of alternative futures for some year involve the same GNP so that meaningful comparisons can be made.

Estimates for 1985 using this model were made using alternatives involving various energy growth rates both with and without the three new technologies. The projections were made with a GNP of $1.34 trillion 1958 dollars for 1985.

Some illustrations of the results are as follows:

- total investment becomes a larger percentage of GNP especially with high BTU gasification, however the other two technologies tested actually decrease investment.
- output of coal mining with coal gasification increases dramatically.
- total employment is nearly constant.

These types of outputs lead to the following types of conclusions:

- total investment in general and capital good industries are quite sensitive to energy use growtl rates.
- major impacts of introducing new energy technology will be on the capital goods industries.
- aggregation of demand for investment funds will occur with introduction of high BTU coal gasification, whereas the gas turbine topping cycle will decrease demand for funds.

The advantages of the input-output approach to evaluating new technologies are apparent from the above conclusions which show that the interactions with all sectors and factors of the economy can be measured. The level of detail is only restrained by the availability of data for the base years. Some of the potential uses for the input-output approach are as follows:

- Evaluating air quality standards time tables in terms of what can be done and at a reasonable cost.
- Evaluating the impacts of multiple investment programs.
- Evaluating the impacts of alternative technological thrusts.

References

[1] "Energy Needs: Projected Demands and How to Reduce Them," *Science*, Vol. 178, 15 December 1972, pp. 1186 to 1188.

[2] D. Chapman, T. Tyrell, T. Mount, "Electricity Demand Growth and the Energy Crisis," Vol. 178, 17 November 1972.

[3] "The Energy Resources of the Earth," M. King Hubbert, *Scientific American*, September 1971.

[4] "Decision Making in the Production of Power," Milton Katz, *Scientific American*, September 1971.

[5] "Operations Research in the Mining Industry," E. M. Ventura, *Progress in Operations Research*, Vol. 2, 1964 John Wiley and Sons, Inc.

[6] "The U.S. Energy Problem," Vols. 1 and 2, Intertechnology Corporation, Warrenton, Virginia 22186.

[7] "Dynamic Energy System Modeling-Interfuel Competition," Martin L. Baughman, August 1972, Energy Analysis and Planning Group, MIT.

[8] "EPA Energy Quality Model Documentation," EPA Internal Document, November 3, 1972.

[9] "California's Electricity Quandary: Estimating Future Demand," W. E. Mooz, C. C. Mow, Rand Corporation, R-1084-NSF CSRA, September 1972.

[10] *The Use of Econometric Models by Federal Regulatory Agencies*, Joe L. Steele, Heath Lexington Books, 1971.

[11] "Bid Evaluation for Procurement of Aviation Fuel at DFSC," H. A. Waggener, G. Suzuki, *Naval Research Logistics Quarterly*, March 1967.

[12] "Techniques Used in Developing the Department of Interior Energy Forecast, United States Energy Through the Year 2000," James A. West, *Energy Modeling*, Resources for the Future Inc. March 1973, pp. 294 to 303.

[13] "NPC Analysis of Oil and Gas Supply," Andrew Avramides and James S. Cross, *Energy Modeling*, Resources for the Future, Inc. March 1973, pp. 304 to 315.

[14] "U.S. Energy Otulook," National Petroleum Council, December 1972.

[15] "A Model for Continental Oil Supply and Distribution," J. G. Debanne, University of Ottawa, *Journal of Petroleum Technology*, September 1971, Vol. 23, pt. 2.

[16] *The Supply and Price of Natural Gas*, M. Adelman, Oxford Press, 1962.

[17] *The Economics of Crude Oil Production*, Paul C. Bradley, North Holland Publishing Company, Amsterdam, 1967.

[18] *Supply and Cost in the United States Petroleum Industry, Two Econometric Studies*, Franklin M. Fisher, Johns Hopkins Press, Baltimore, Maryland, 1964.

[19] "The FPC Staff's Econometric Model of Natural Gas Supply in the United States," D. J. Khazzoom, *The Bell Journal of Economics and Manage-*

ment Science, Vol. 2, No. 1, Spring, 1971.

[20] *Competition in the ~~Midwestern~~ Coal Industry*, Moyer Reed, Harvard University Press, Cambridge, Massachusetts, 1964

[21] *The Demand for Natural Gas in the United States*, Pietro Balestra, Harvard University Press, Cambridge, Massachusetts, 1966.

[22] *A Study in Econometrics: The Demand for Electricity in the United States*, Franklin Fisher, North Holland Publishing Company, Amsterdam, 1962.

[23] "Environmental Screening Model for the Location of Power Generating Facilities," David H. Marks, Gerhard H. Jirka, Paper presented at the 39th Meeting of the Operations Research Society of America, May 7, 1971.

[24] "Mathematical Modeling in the Maryland Power Plant Monitoring Program," Werner Furth, Corporate Research and Development Laboratory of Martin Marietta, Internal Paper.

[25] "Interfuel Competition for Future Energy Markets," Richard Gonzales, *Journal of the Institute of Petroleum*, Vol. 54, No. 535, July 1968.

[26] Status and Consequences of Urban Heat Rejection," Cornell Energy Project, Kenneth Torrance, February 1972.

[27] "A Review and Comparison of Selected U.S. Energy Forecasts," OST, Energy Policy Staff, Battelle Memorial Institute, December 1969.

[28] "Technology Assessment with Special Reference to Energy," C. L. Comar, Cornell Energy Project, October 1970.

[29] "Electric Utility Optimum Mix Model," K. B. Cady, J. Hass, Cornell Energy Project, 1971.

[30] "Computer Simulation in Open Pits," R. G. Miller, RT2, Consultants Ltd., London, Energy Mining Magazine, September 1972.

[31] "Interfuel Substitutability in the Electric Utility Sector of the U.S. Economy," Thomas Duchesneau, Economic Report to the Federal Trade Commission, February 1972.

[32] "Public Utility Investment Analysis," J. Hass, January 1971, Cornell Energy Project.

[33] "Poverty and Economic Growth," J. E. Hass, G. Soutar, March 1972, Cornell Energy Project.

[34] "Electric Space Heating," Tom Bolan, June 1971, Cornell Energy Project.

[35] "Development of Linear Programming Total Energy Model," Atomic Energy Commission Report.

[36] "California's Electricity Quandry: Planning for Power Plant Citing," R. H. Ball, R. J. Sartre, et. al., Rand Report, R-1115-RF/CSA, September 1972.

[37] "California's Electricity Quandry: Slowing the Growth Rate," R. D. Doctor, K. P. Anderson, et. al., Rand Report, R-1116-NSF/CSA, September 1972.

[38] "Fuels for the Electric Utility Industry: 1971 to 1985, NERA Economic Consulting Staff, National Economic Research Associates, Inc., August 15, 1972.

[39] "Patterns of Energy Consumption in the United States," OST, U. S. Government Printing Office #4106-0034.

[40] "An Energy Model for the United States," Bureau of Mines, Report No. IC8384.

[41] "Electricity Demand Growth Implications for Research and Development," D. Chapman and T. Mount, Oak Ridge National Laboratory, August 1972.

[42] "Toward Econometric Estimation of Industrial Energy Demand, Coal, and Experimental Application to the Primary Metals Industry," K. P. Anderson, Rand Report, R-719-NSF, December 1971.

[43] "A Review and Comparison of Selected United States Energy Forecasts," Battelle Memorial Institute, Columbus, Ohio, December 1969.

[44] "Quantitative Models in the Energy Sector: A Review of the State of the Art," J. R. Sharko, D. R. Limaye, Decision Sciences Corporation, Report No. 114.

[45] "TERA, A Total Energy Resources Analysis Model," D. R. Limaye, Decision Sciences Corporation.

[46] "Optimization of the North American Crude Oil Distribution System," P. R. Gosselin, Operations Research Branch, National Energy Board, Ottawa, Canada.

[47] "A Three Dimensionable Model for Predicting Natural Gas Reserves and Availability," Ross C. Richards, National Energy Board, Ottawa, Canada.

[48] "Forecasting Techniques with Some Applications to Non-Renewable Energy Resources," G. T. McLoughlin, June 8, 1972, National Energy Board of Canada, Operations Research Branch.

[49] "Quantitative Energy Studies and Models, A State of the Art Review," Decision Sciences Corporation, prepared for Council on Environmental Quality, undated.

[50] "Input-Output Analysis of U.S. Energy Consumption," William A. Reardon, *Energy Modeling*, Resources for the Future, Inc. March 1973, pp. 23 to 44.

[51] "Use of the Maryland Interindustry Forecasting Model to Project Petroleum Demand," Clopper Almon, *Energy Modeling* Resources for the Future, Inc. March 1973, pp. 45 to 61.

[52] "An Econometric Analysis of the Relationships between Macro Economic Activity and U.S. Energy Consumption," Philip K. Verleger, Jr., *Energy Modeling*, Resources for the Future, Inc., March 1973, pp. 62 to 102.

[53] "A Unified Framework for Energy System Planning," Kenneth C. Hoffman, *Energy Modeling*, Resources for the Future, Inc., March 1973, pp. 108 to 143.

[54] "Linear Programming in Energy Modeling," D. E. Deonigi and R. L. Engel, *Energy Modeling*, Resources for the Future, Inc., March 1973, pp. 144 to 176.

[55] "Suggested Roles for Econometrics and Process Analysis in Long Term Energy Modeling," James M. Griffin, *Energy Modeling*, Resources for the Future, Inc., March 1973, pp. 177 to 185.

[56] "Substitution and Usage in Energy Demand: An Econometric Estimation of Long Run and Short Run Effects," Edward W. Erickson, Robert M. Spann and Robert Ciliano, *Energy Modeling*, Resources for the Future, Inc., March 1973, pp. 190 to 208.

[57] "Joint Costs and Separability in Oil and Gas Exploration," Robert M. Spann and Edward W. Erickson, *Energy Modeling*, March 1973, Resources for the Future, Inc., pp. 251 to 265.

[58] *Implication of the Growth and Demand for Commercial and Industrial Electrical Energy in the South Coast Air Basin* Lester Lees, EQL Report #2 Environmental Quality Laboratory, California Institute of Technology, November 1971.

[59] *Energy Use in California: Implication for the Environment*, E. J. List, Environmental Quality Laboratory, California Institute of Technology, EQL Report No. 3, December 1971.

[60] *Fuels for the Electric Utility Industry: 1971 to 1985*, NERA Economic Consulting Staff, National Economic Research Associates, Inc., New York, New York, August 15, 1972.

[61] *A National Energy Model*, Cecil H. Chilton, Battelle Memorial Institute, Columbus, Ohio.

[62] *Reference Energy Systems and Resource Data for use in the Assessment of Energy Technologies*; Report No. AET-8, Associated Universities, Inc., April 1972.

[63] *Forecast of Growth of Nuclear Power*, January 1971, U. S. Atomic Energy Commission, Division of Operations Analysis and Forecasting.

[64] *Natural Gas Supply and Demand, 1971 to 1990*, Staff Report No. 2, Bureau of Natural Gas, Federal Power Commission, Washington, D.C., February 1972.

[65] *Energy Policy Issues for the U.S. During the Seventies*, Prepared for the National Energy Forum, sponsored by the National Division, U.S. National Committee, World Energy Conference, Arthur D. Little, Inc.

[66] *The Availability and Requirements of Stationary Source Fossile Fuels, 1972 and 1977*, Mitre, MTR-6221, August 1972.

[67] *Energy Consumption by States. Historical Trends and Projections from 1975 to 1980*, Helmut J. Frank, Jean Webber and David Monarchi Division of Economic and Business Research, College of Business and Public Administration, University of Arizona, Tucson, Arizona, June 1972.

[68] *Consumption of Electricity in the United States* by Oran L. Culbertson, Oak Ridge National Laboratories, Oak Ridge Tenn. Report No. ORNL No. NSF-EP-5.

[69] *Utilization Analysis of Energy Systems; Fundamentals of Energy Utilization*, by H. L. Brown and B. B. Hamel; and *Application of Utilization Criteria to House Energy Systems*, by B. B. Hamel, H. L. Brown, R. E. Crain, and J. K. Mwude, Energy Sources Systems Institute, Drexel University, Philadelphia, Pa. December 1971.

[70] *Supply and Demand for Energy in the United States by States and Regions 1960 to 1965*, Bureau of Mines Information Circular 8434, Bureau of Mines, United States Department of the Interior, 1970.

[71] *Survey of Energy Consumption Projections*, Committee on Interior and Insular Affairs, U.S. Senate pursuant to Senate Resolution 45, A National Fuels and Energy Policy Study, Serial No. 92-19, U.S. Government Printing Office, Washington, D.C., 1972.

[72] *Outlook for Energy in the United States to 1985*, by J. G. Winger, et al. Energy Economics Division, Chase Manhattan Bank, New York, New York, June 1972.

[73] *Power Pollution and Public Policy*, Dennis Ducsile, M.I.T. Press, 1971.

[74] *Impacts of New Energy Technology Using Generalized Input-Output Analysis*, James E. Just, Mitre Corporation, Report M73-32, February 1973.

[75] *Changes in Energy Consumption 1963-1980*, James E. Just, MS thesis, MIT, February 1973.

Chapter 11

Models in Planning and Operating Health Services

By

Boyd Z. Palmer

SUMMARY 349
I. THE HEALTH SERVICES SYSTEM 349
  Introduction 349
  Health Services as a System 351
  Performance Measures 353
II. NATIONAL POLICY MODELS 354
  National Policy 354
  Population Growth Policy 355
  Health Services Supply 355
  Delivery Systems Design 359
  National Program Evaluation and Comparison 360
III. COMPREHENSIVE HEALTH PLANNING MODELS 361
  Regional Health Planning Decisions 361
  Adaptive Control Models 361
  Probability Models 362
  Utilization Simulation Models 363
  Econometric Models 365
  Mathematical Programming Model 365
  Ill-Health Generation Models 366
IV. COMMUNITY HEALTH SERVICE ALTERNATIVES 366
  Community Health Service Decisions 366
  Blood Bank Models 366
  Emergency Medical Care Models 367
  Planning Alcoholism Programs 367
  Maternal and Infant Care Programs 367
  Family Planning Programs 367
  Narcotics Control Models 367
V. INSTITUTIONAL PLANNING AND CONTROL MODELS 368
  Facility Planning and Control Decisions 368
VI. ASSESSMENT OF HEALTH SYSTEM MODELS FOR POLICY ANALYSIS 368
  Model Types Versus Level of Decision 368
  Prognosis and Conclusion 369
REFERENCES 370

Preceding page blank

Models in Planning and Operating Health Services

SUMMARY

This chapter reviews health services models and concludes with a general assessment of these models from the viewpoint of decision-makers in health planning or in health service organizations. The discussions emphasize the modeling approaches taken, their development stage and the performance measures used, rather than specific technical details of model construction or data estimation. The basic question to be examined is: "to what extent have models been or can be used to estimate future effects of alternative proposals for improving health."

The chapter begins by reviewing difficulties of developing useful models of health services and the general approaches that modelers have taken to overcome these problems. Major problems noted include the lack of agreed-on measures of health status, the lack of established relationships between resource use and health conditions, and the large number of dimensions that must be built into any realistic representation of health services.

The applications of models are reviewed in three areas: Federal policy decisions; regional, state and area-wide planning decisions; and community health service alternatives. Models are also useful for institutional planning and control decisions and this operational use is surveyed briefly. These areas do not exhaust the application of models to health problems (for example, there are many models of the nervous system and the circulatory system, diets, and biomedical research), but they should serve to give an overview of the possibilities and problems of the use of models to analyze health system decisions with scope ranging from community to national.

The Summary Table lists the more important models discussed in this chapter.

I. THE HEALTH SERVICES SYSTEM

Introduction

Models of health services date back to the early 1950's in the U. S. and perhaps somewhat earlier in Great Britain. Those early efforts were primarily concerned with the application of probability concepts to relatively small portions of a particular health service, such as a radiology department of a hospital. More recently, several factors have led to great interest in analytic techniques that might help explain what happened as a result of past actions, e.g. the introduction of Medicare, and predict what might happen if proposed actions were to be taken, e.g. national health insurance.

The need to analyze health systems results from the emergence of health services as a major, fast-growing sector of the U. S. economy, the realization that large groups of people receive inadequate care because of their inability to purchase services, and the realization that there is uncertainty as to even first-order effects of Federal actions to alleviate maldistribution problems.

Several models have been developed to estimate, on a national scale, effects of Federal policy actions affecting supply of hospitals, training of health manpower, changing the organization of the health service delivery system, and

Preceding page blank

HEALTH SERVICES

Models Discussed in Chapter

Summary Table

Model Decision Area	General Type	Important Characteristics
National Policy Models:		
Population Growth	Dynamic feedback	Analysis of long-term impact of population growth and health policies
Health Service	Macroeconomic Microeconomic Simulation	Evaluation of manpower programs, medicare, credentialling rules, hospital services required for population and demand alternatives
Health Services Demand	Statistical	Estimate of demand for hospital care, physician, etc.
Delivery Systems	Mathematical programming Simulation	Evaluation of HMO's, alternative health delivery organizations
Program Evaluation and Comparison	Statistical Mathematical programming	Analysis of disease control programs
Comprehensive Health Planning:		
Regional Health Planning	Adaptive control Probability Simulation	Analysis of alternative actions, health goals and objectives, estimate future resource requirements
Demand and Supply	Analytic Deterministic	Estimate demand and supply of health services and measure system performance
Health Services Impact	Statistical Simulation Econometric	Evaluation of how a health services system reacts to major shocks, e.g. disasters
Utilization of Services	Simulation	Forecast of resource requirements
Vancouver and Quebec Health Care System	Simulation	Study of health care system within total social environment, evaluation of alternative resource allocations
Comprehensive Health Planning Agency	Econometric	Use by agency to forecast and evaluate policies
Minnesota Health Planning Agencies	Econometric	Use by agencies to forecast and evaluate policies
Mental Health Services	Mathematical programming	Analysis of alternative policies
Ill-Health Generation	Simulation Statistical	Estimate of air pollution medical effects, health service needs
Community Health Service:		
Blood Bank	Inventory Simulation Statistical	Evaluation of management policies, forecasting requirements, allocation of supplies
Emergency Medical	Simulation Probability	Determine the location and number of ambulances
Alcoholism Program	Simulation	Evaluation of alternative programs
Maternal and Infant Care Program	Econometric Simulation	Evaluation of effects of changes in population or nursing staffs
Family Planning Programs	Simulation	Determine patient flow requirements for budget estimating
Narcotic Control	Probabilistic Simulation	Cost benefit analysis of alternative programs, analysis of the economics of control programs

Institutional Planning and Control:
(No detail discussion given—see Section V for range of problems treated)

enabling people to buy more services through greater insurance coverage.

The advent of a legislative requirement for comprehensive health planning at the state and regional level has led to the development of regional health models. At the same time, modeling has continued as a technique to aid institutional management, particularly hospitals, to assist in the evaluation and comparison of programs such as blood banking, emergency medical services, alcoholism, and tuberculosis treatment.

This chapter describes some of the more important models developed to analyze decision problems which arise in the health services area. Before proceeding to example models, we first give an overview of the health services system.

Health Services as a System

In order to appreciate the difficulty in developing models which depend on visualizing health services as a system, it is necessary to understand the complexity of health services interactions, and the lack of general agreement on definitions and classifications.

Health services are now provided by a variety of institutions, independent practitioners, governmental and private agencies. The interactions among these agencies are very loose, being connected only by those processes which connect any institutions within a society: its economic markets, overall governmental (executive and legislative) processes, and normal methods of coordinating human activities. There is not, in any region, a central health services delivery, management or planning authority. Thus, health services are characterized by most health professionals as a "non-system." The separate services are regarded as autonomous, and the service providers have simply evolved as the result of a variety of forces, internal and external, acting upon them. The present "non-system" is so complex that it even involves agencies with dual authorities, the prime example being the separation of administration and medical direction in a hospital. Each institutional provider of health service can make its own decisions on which persons to accept as patients, and which to refer to other providers, without follow-up to see if the referred people actually are accepted by any other provider. Each provider can create his own data system, using different classifications. Mental health services are completely separate from physical and dental services, and personal health services are considered to be separate from environmental health services. There are different entry points and procedures among providers, depending to a large extent on location, e.g., urban, suburban, rural; and on the financial resources of the consumer available for health maintenance.

These characteristics of health services have posed formidable problems to health system modelers, because many potentially relevant model types depend on defining the jumble of services as a system, not in the sense that a central seat of power exists that can control (or manage) all services in the system, but in the sense that health services and the flows of people from one service to another, can be defined, measured, and analyzed. If health services can be assumed to be components of a system, traditional systems concepts of input, output, capacity, processing, and system performance can then be utilized, and a model constructed based on application of these concepts to the decision area to be analyzed. A further problem confronting health modelers is that suitable measures of health service system performance and data consistent with a systems representation cannot be obtained. Pervading the whole analysis effort is the dimensionality of the health area which necessitates aggregation in the major variables, and disregard of many relevant variables, if the effort is to be at all feasible.

Figure 1 illustrates the complexity and dimensionality problems using the common approach that health services should be evaluated fundamentally by

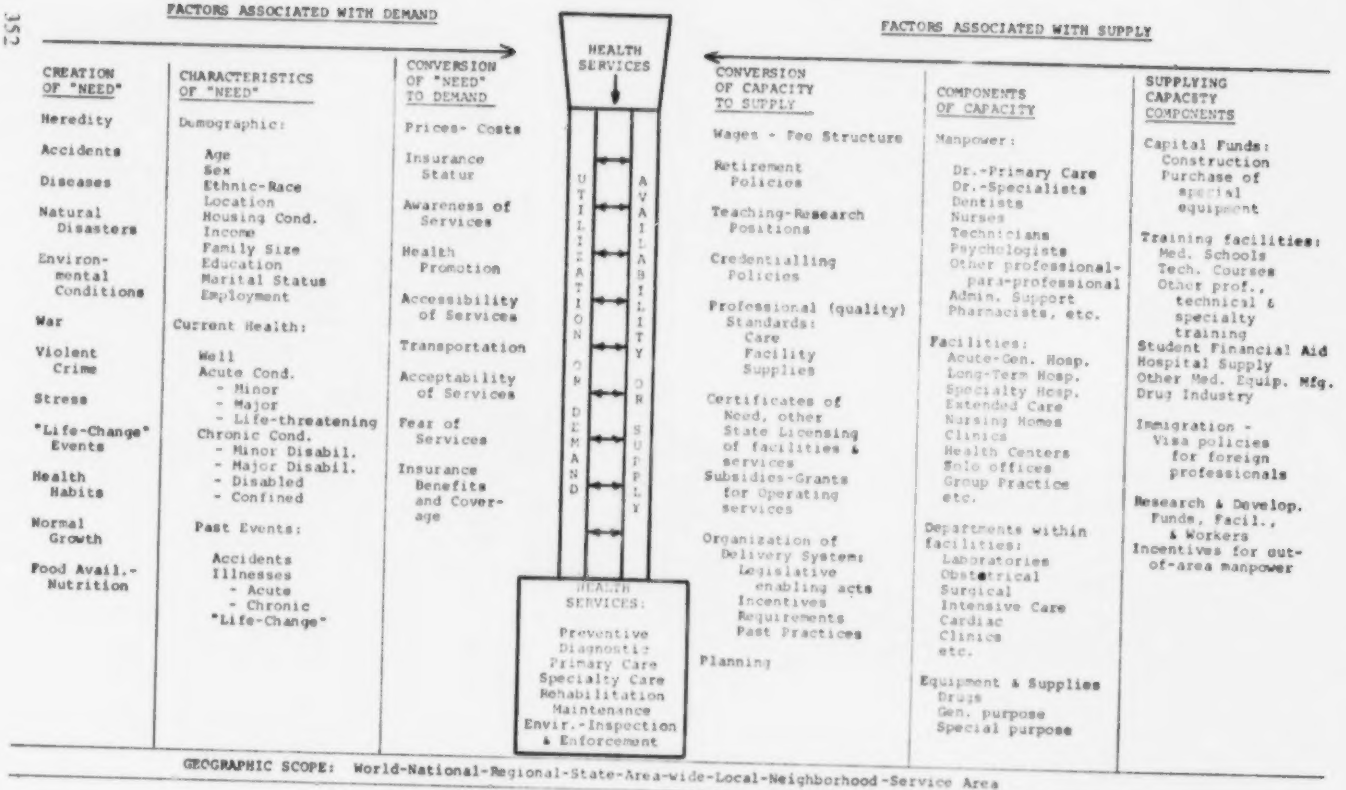

FACTORS ASSOCIATED WITH DEMAND
CREATION OF "NEED"
Heredity
Accidents
Diseases
Natural Disasters
Environ-mental Conditions
War
Violent Crime
Stress
"Life-Change" Events
Health Habits
Normal Growth
Food Avail.-Nutrition
CHARACTERISTICS OF "NEED"
Demographic:
Age
Sex
Ethnic-Race
Location
Housing Cond.
Income
Family Size
Education
Marital Status
Employment
Current Health:
Well
Acute Cond.
- Minor
- Major
- Life-threatening
Chronic Cond.
- Minor Disabil.
- Major Disabil.
- Disabled
- Confined
Past Events:
Accidents
Illnesses
- Acute
- Chronic
"Life-Change"
CONVERSION OF "NEED" TO DEMAND
Prices- Costs
Insurance Status
Awareness of Services
Health Promotion
Accessibility of Services
Transportation
Acceptability of Services
Fear of Services
Insurance Benefits and Coverage
HEALTH SERVICES
UTILIZATION OR DEMAND
AVAILABILITY OR SUPPLY
HEALTH SERVICES:
Preventive
Diagnostic
Primary Care
Specialty Care
Rehabilitation
Maintenance
Envir.-Inspection & Enforcement
FACTORS ASSOCIATED WITH SUPPLY
CONVERSION OF CAPACITY TO SUPPLY
Wages - Fee Structure
Retirement Policies
Teaching-Research Positions
Credentialling Policies
Professional (quality) Standards:
Care
Facility
Supplies
Certificates of Need, other State Licensing of facilities & services
Subsidies-Grants for Operating services
Organization of Delivery System:
Legislative enabling acts
Incentives
Requirements
Past Practices
Planning
COMPONENTS OF CAPACITY
Manpower:
Dr.-Primary Care
Dr.-Specialists
Dentists
Nurses
Technicians
Psychologists
Other professional-para-professional
Admin. Support
Pharmacists, etc.
Facilities:
Acute-Gen. Hosp.
Long-Term Hosp.
Specialty Hosp.
Extended Care
Nursing Homes
Clinics
Health Centers
Solo offices
Group Practice
etc.
Departments within facilities:
Laboratories
Obstetrical
Surgical
Intensive Care
Cardiac
Clinics
etc.
Equipment & Supplies
Drugs
Gen. purpose
Special purpose
SUPPLYING CAPACITY COMPONENTS
Capital Funds:
Construction
Purchase of special equipment
Training facilities:
Med. Schools
Tech. Courses
Other prof., technical & specialty training
Student Financial Aid
Hospital Supply
Other Med. Equip. Mfg.
Drug Industry
Immigration -
Visa policies for foreign professionals
Research & Develop.
Funds, Facil., & Workers
Incentives for out-of-area manpower
GEOGRAPHIC SCOPE: World-National-Regional-State-Area-wide-Local-Neighborhood-Service Area

FIGURE 1—*Modeling Health Services: Potential Areas and Factors*

comparing demand (or need) with supply (or capacity). Assumptions incorporated into this figure are: (1) The basis for relating supply and demand is via services, the processes of providing health care; (2) The health services system should include physical, dental, mental, and environmental health service activities; (3) The need for services is created by the processes of living which are outside the health services system; (4) In general, need is not the same as demand; the latter is created when the perception of the need is combined with recognition of an available, accessible source of help (service); (5) The capacity for health services must be created by providing the proper combination of basic resources; (6) Even the smallest geographic scope of interest (a neighborhood) can include virtually the entire range of factors associated with demand for, and supply of health services.

Most modeling efforts which attempt to represent a broad range of services (rather than institution specific models) can be located on Figure 1 by specifying geographic scope and the demand and supply factors. Which particular factors are selected for representation in a given model depends, of course, on the purposes of the modeling effort and the decisions to be analyzed.

In general the existing health models have dealt with the problems of complexity, lack of standard measures, dimensionality, and lack of data, in the following ways:

- through use of a "non-system" model (i.e. statistical regression model), which does not require consistent measures of input, output, resources, and performance.
- through use of a system (formal) model, with relatively few highly aggregated variables, and using hypothetical data values.
- by narrowing the scope, in terms of services covered and/or geographic area, to a feasible system which can then be modeled with available, consistent data.
- by designing but not constructing a detailed broad scope model to show how problems could be analyzed if adequate data were available and measures used were acceptable.

All models do not fit neatly into these groups; many are hybrids of systems and non-systems models, and many use a scope just a little broader than available data will fit, thus requiring reasonable estimates of some of the data. Just about all models narrow the scope to either physical health services, excluding mental and environmental health services, or mental health services alone, or environmental services alone. Measurements to link service utilization and resultant changes in health conditions do not exist, so system performance measures have to be defined in other ways. In fact, the variety of definitions of performance used in models within the same decision area is a major factor affecting the usefulness of the models for decision analysis, and greatly hampers comparison of models for such purposes. As measurement of health services activities is of such critical importance to the policy maker, as well as the model maker, we next review some approaches to performance measures.

Performance Measures

These can be grouped into two general types: *outcome* measures, which concentrate on health status of people using health services, and *process* measures, which represent various characteristics of the service process; we discuss examples of each type in turn.

Data series which are used to measure health status of individuals are usually built on states of disability. Garfield [36] proposed a classification of well, worried well, early sick, sick, and very sick categories, so health services appropriate to each category could be assigned.

Torrance, et al. [110], conceptualized a three dimensional disability state measure, based on physical function, emotional function and social function.

A linear interval scale was proposed with a value of 1 being equivalent to "perfect health" in all functions, and a value of 0 being equivalent to death. Rather than define all disability states, an index value based on this scale is suggested. This utility index was used, with hypothetical data, in a mathematical programming formulation of a health program ranking procedure.

Chorba, Miller, and Sanders [14] suggest a "deprivation" index based on time spent seeking health care, time spent in bed due to sickness, and time lost due to death before age 90. This index was used in a model analyzing tuberculosis program alternatives [15]. Blum [6] suggested a similar index for use in assigning priorities to programs. Fanshel and Bush [26] derived a dysfunction index based on ability to function in society. This work has been further refined by Bush, et. al. [8 and 9] in a Health Index Project at the University of California. Thirty status categories have been defined for evaluating success of health programs. Success means moving people into better status categories. This approach results in a single-valued health index; it has been used in a model to evaluate tuberculin testing programs [11].

Many variations of morbidity and mortality measures have also been used. So far, none of the proposed health status measures have received the prestige of official approval by professional organizations nor general usage by a large number of health researchers.

A common procedure in health modeling is to use process measures, i.e. calculate demands for services and supply of those services, and estimate "unmet demands" by the difference. The implied assumption is that all persons who received services needed the services and such persons gained improved health status as a result. Thus, process-oriented performance measures emphasize utilization of services, rather than improvement in health.

At this time, models which use process-oriented performance measures or none at all are predominant among health models. They represent a feasible compromise between the desire to build models that can optimize health levels using accepted measures of health, and the known limitations of data now available.

II. NATIONAL POLICY MODELS

National Policy

The economy of the United States has traditionally operated on a "market" basis which incorporates the concept of price mechanisms. In the area of health, these mechanisms mean that some people are unable to buy services in the main delivery system. A series of government actions has attempted to help people purchase services, or to set up alternative delivery systems. The Federal government has in the past chosen to allocate an increasing amount of its resources to increase the overall national supply of hospitals, mental health centers, and physicians. It has provided subsidies to encourage reallocating resources and services to the elderly, the young, and the disadvantaged, both in the core city and in isolated rural areas. As a result of continuing problems and disagreement about the success of prior actions, basic questions have arisen about the proper role for the Federal government in health. Some of the important questions are: Should the Federal government supply funds to build hospital facilities? Should it support categorical disease-control programs directly or by revenue-sharing? Should it begin a national health insurance program? If so, should such insurance have limited or comprehensive coverage? Should Health Maintenance Organizations be encouraged? If so, what type of organization would give best health care? Should the Federal government actually operate any health services directly, or simply supply incentives to affect market forces? Should basic research in health services be increased or decreased?

Many alternative answers to these questions have been suggested [35,58, 63,102,113]. Programs adopted in the next few years will guide actions, funding, legislation, and regulations for some time to come. We next discuss models which address some of these questions or which have a bearing in determining national health programs.

Population Growth Policy

The broadest health policy question to be supported by models is what should be the rate of population growth to produce the "best" health of the population. Meadows, et al. [67] constructed a world model which included one aggregate measure of health services and its delayed impact on life expectancy. This model contains feedback processes, has only a few, highly aggregated variables and uses estimates of parameters. It computes values for important variables many years into the future. It is not intended for detailed policy analysis; its main use is to indicate the need for population growth policies (among others) by demonstrating the general impact of alternative values of the key parameters.

A similar type of model was used by Kane, Thompson and Vertinsky [56], to investigate aggregate effects over 25 years of three Canadian health policy intervention strategies on population size, mortality rate, availability (capacity) of health services, total health expenditures, level of diagnostic screening and the delegation of physician functions.

The three intervention strategies used were:

Policy A—"Fire-fighting." The model assumed incremental adjustments in health service whenever morbidity got too high, expenditures got too high, or availability got too low.

Policy B—Maintain high availability of health service. Two methods were used: (1) keep population at zero growth, and (2) increase facilities and delegate more physician functions to paramedics.

Policy C—Increase screening to detect health problems. Screening was tripled over the first five year period, with no further intervention thereafter.

Policies B and C amount to major restructuring of the health services system. They are general policies which would be realized by a multitude of specific actions, each of which could be done in alternative ways.

The model results showed that Policy A was a "clear failure, giving very little in return for its added expense," and the others were also regarded as too expensive. The only solution that might avoid future financial crises would be to alter the productivity of medical resources.

Health Services Supply

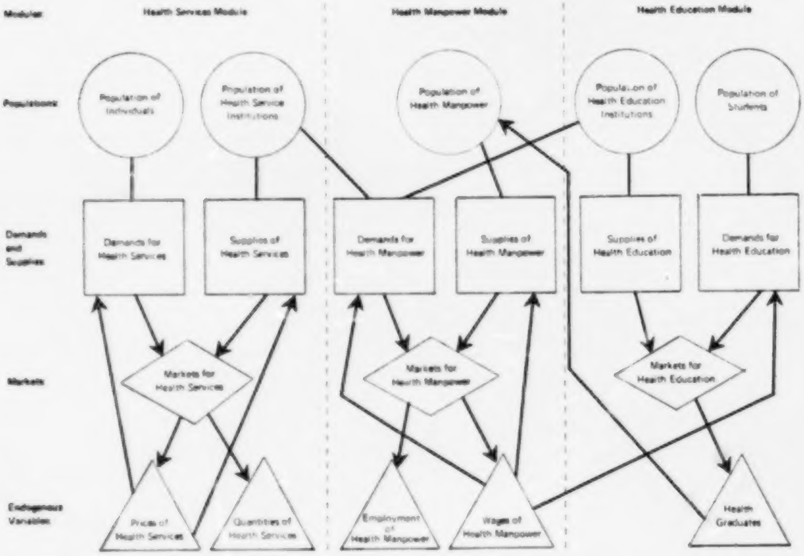
The question which has attracted the most attention from health service researchers deals with the national supply of health facilities and health manpower: what is the current gap between supply and demand, and how will the gap change in the future? The most common model form used to examine such questions has been aggregated, macroeconomic models, consisting of a large number of equations designed to explain variables expressing supply of medical manpower and hospital beds, utilization of physician time and hospital beds, and some prices, costs, and government expenditures. Generally, they estimate effects of past policy changes and explain what has already happened. Feldstein [27] developed such a model to study the impact of Medicare and make projections of possible future changes; see also [28] and [94].

In 1969, the U. S. Bureau of Health Manpower Education contracted for pilot studies in the development of microsimulation models (that is, detailed simulations) for health manpower estimation. During the summer of 1970, a two day conference was held to present and discuss results of the designs. The two-volume proceedings of that conference constitutes a basic reference for health manpower modeling, and on microsimulation models in general.

Volume One [121] presents an *econometric* microsimulation approach which was investigated by the Human Resources Research Center of the University of Southern California. This group also developed macroeconomic models of the national health system (Yett, et al. [124,125]). The research effort conceptualized two highly complex models, called Mark I and Mark II. While most econometric models deal with only one "population" (such as health manpower), both I and II use five populations: individuals, health service institutions, health manpower, health professions educational institutions, and students. Using three computational modules, for health services, health manpower, and health education, the five populations interact in four basic types of markets: for outpatient services, inpatient services, manpower, and training. This is diagrammed in Figure 2. Mark II is a scaled down version of I, but still involves hundreds of equations.

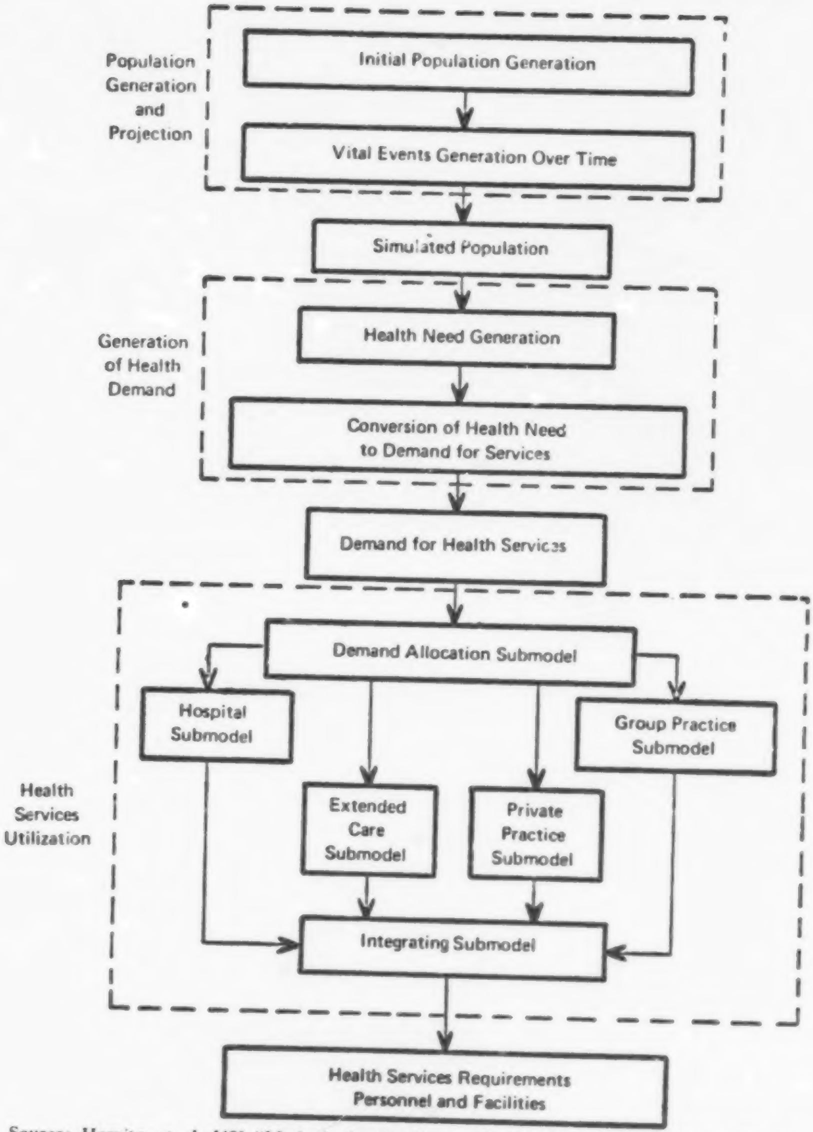
The proposed use of these models would be to concentrate on manpower issues although other policy issues could also be evaluated. The report discusses ways in which the Mark I model could be used to test changes in Medicare coverage, introduction of new health occupations (e.g. physician assistant), and changes in credentialling rules (e.g. education requirements in nursing programs). Neither model can handle issues about geographic maldistribution of services, and neither includes dental, mental or environmental services.

Volume Two of the proceedings presents a probabilistic microsimulation model developed by the Research Triangle Institute (RTI) [48]. RTI was required to produce a working model, using whatever data they could find. This led the RTI researchers to concentrate on the hospital portion of the health system. Figure 3 shows a diagram of the more complete model de-


Modules
Health Services Module
Health Manpower Module
Health Education Module
Populations
Population of Individuals
Population of Health Service Institutions
Population of Health Manpower
Population of Health Education Institutions
Population of Students
Demands and Supplies
Demands for Health Services
Supplies of Health Services
Demands for Health Manpower
Supplies of Health Manpower
Supplies of Health Education
Demands for Health Education
Markets
Markets for Health Services
Markets for Health Manpower
Markets for Health Education
Endogenous Variables
Prices of Health Services
Quantities of Health Services
Employment of Health Manpower
Wages of Health Manpower
Health Graduates

Source: Yett, et al. [121] "Conceptualization of a Health Manpower Simulation Model"; *Vol. I of Proceedings of Conference on a Health Manpower Simulation Model*, HEW, NIH, Dec. '70; Page 29.

FIGURE 2—*Basic Structure of the USC Micro-Simulation Model*


Population Generation and Projection
Initial Population Generation
Vital Events Generation Over Time
Simulated Population
Generation of Health Demand
Health Need Generation
Conversion of Health Need to Demand for Services
Demand for Health Services
Health Services Utilization
Demand Allocation Submodel
Hospital Submodel
Extended Care Submodel
Private Practice Submodel
Group Practice Submodel
Integrating Submodel
Health Services Requirements Personnel and Facilities

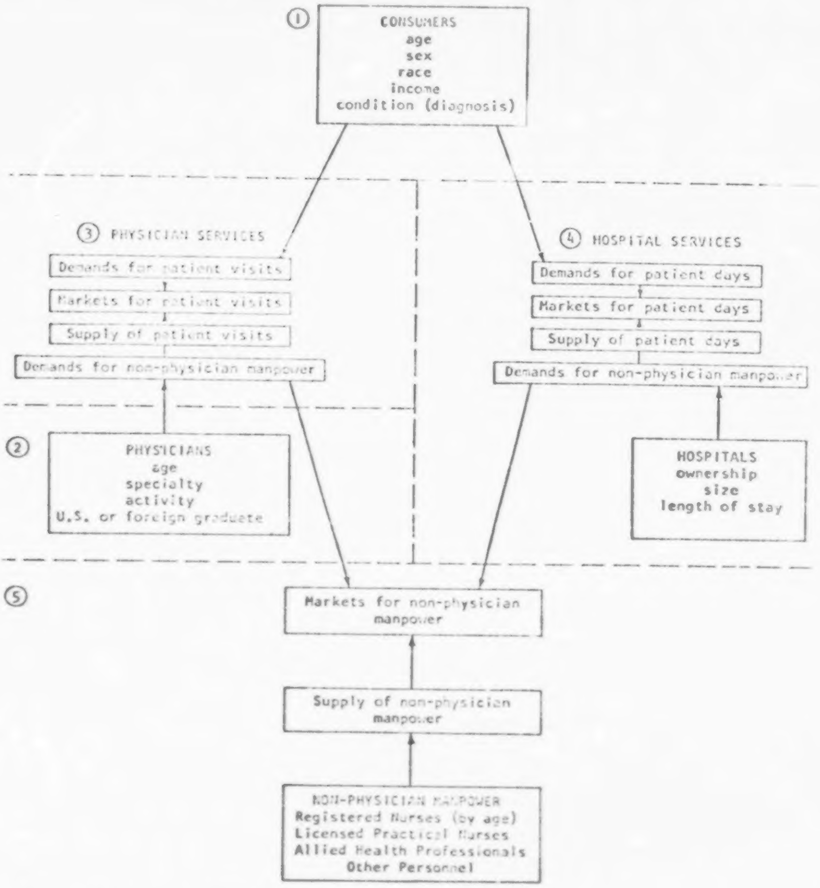
Source: Horvitz, et al. [48] "Methods for Building a Health Manpower Simulation Model", *Vol. 2: Proceedings of Conference on A Health Manpower Simulation Model*, HEW, NIH, Dec. '70, page 49.

FIGURE 3—*Health Care Demand System Model*

sign rather than the actual programmed model which considered only hospital services and hospital manpower. The uses of the model centered on measuring variations in requirements for hospital services when population distributions and demand estimates changed.

The differences between the micro and macro approaches, at least using econometric models, is discussed in a relatively non-technical article in Yett, et. al. [122], which compares the Mark I model with a macro model for regional planning.

A continuation of the above efforts developed a scaled down, but operational version of the Mark II model, Yett, et. al. [123]. This represented a significant achievement in data assembly and econometric analysis and has the potential to be a useful econometric model of the health services system. While the model has not as yet been used for prediction purposes, the current version demonstrates that it is possible to obtain an integrated working representation of services, resources supply, consumer demands, and market prices, within the constraints of available data. Figure 4 shows a block


① CONSUMERS
age
sex
race
income
condition (diagnosis)
③ PHYSICIAN SERVICES
Demands for patient visits
Markets for patient visits
Supply of patient visits
Demands for non-physician manpower
④ HOSPITAL SERVICES
Demands for patient days
Markets for patient days
Supply of patient days
Demands for non-physician manpower
② PHYSICIANS
age
specialty
activity
U.S. or foreign graduate
HOSPITALS
ownership
size
length of stay
⑤ Markets for non-physician manpower
Supply of non-physician manpower
NON-PHYSICIAN MANPOWER
Registered Nurses (by age)
Licensed Practical Nurses
Allied Health Professionals
Other Personnel

Source: Yett, et al., [123] *"The Preliminary Operational HRRC Model"* Human Resources Research Center, U. of So. Cal., 1973.

FIGURE 4—*Block Diagram of the HRRC Microsimulation Model*

diagram of this model, which is composed of 5 sub-models.

The first submodel generates a population of consumers or individuals who demand medical services. It estimates the nation's annual population subdivided into cells according to the attributes age, sex, race, income, and condition or diagnosis. The three major events which change the overall population are birth, death, and immigration. The second submodel generates a population of physicians, providing annual estimates of the supply of doctors, subdivided into cells according to their age, specialty, and type of professional activity. The physician services or third submodel computes: (i) the demands (by each consumer group) for patient visits with physicans in private practice and on hospital-based clinics; (ii) the supply of patient visit capacity provided by doctors in office-based practice; and (iii) the doctors demands for aides (i.e., non-physician manpower). The fourth submodel computes the demands for patient days at short-term and long-term hospitals, and the fifth submodel computes the supply of non-physician manpower. The solution procedure balances the variables among the submodels for each period (year) of the simulation.

Many other econometric models relating to health services have been developed. A review of the economic factors used in projecting requirements for health manpower is given in Klarman [59]; Yett, et. al. [125] contains a comprehensive review of economic research studies covering many aspects of the health service system. However, most of these studies are difficult for non-economists to understand, and do not directly relate to public policy decisions. For example, Grossman [40] investigated the demand for health, rather than health services, and used 1963 survey data from the Chicago based National Opinion Research Center to estimate coefficients of his econometric equations. He treats good health as a commodity; specifically, as a "durable good" that depreciates over time, but can be increased by investment in medical care. The analysis can then compare investments in health care in relation to other investment opportunities, such as in housing, recreation, and good. The purpose of the study was to develop a theoretical model and test it with empirical data. No predictions related to public policy were contemplated. But the understanding gained about the relationships between health levels and medical expenditures may prove of value to future work in econometric modeling of health problems.

Non-economic approaches to estimation of health services have also been developed. Navarro and Joroff [76] studied the distribution of physicians in the urban areas of the U. S., using statistical multi-variate analysis. Wirick [118] used a statistical model to estimate demand for hospital days of care, physician visits, dental visits, expenditures for prescribed medicines, and expenditures for other medical expenses.

Finally, the impact of changes in several Federal political policies in hospitals was investigated statistically by Jaeger [51]. Availability, accessibility, cost, financing, manpower, efficiency, and quality measures were used as dependent variables, with fiscal policy, discrimination "policy," population growth, education, and economic ability used as independent variables. The concept of the model, and the measures used, are of interest; numerical results proved inconclusive and of little value for policy makers.

Delivery Systems Design

There are strong indications that the real problem of health care in the U. S. is not the demand-supply relationship (which might be stabilized by increasing capacity or even by decreasing demand), but the way in which health care is organized. For example, Glazer [37] and Joyce [55] examine paradoxical evidence which indicates that there may already be a sufficient supply of health care. The evidence, based on wartime depletion of medical manpower, shows that a direct relationship between supply of physicians and mortality rates cannot be established. Garfield [36] examines health service utilization differences between pre-paid care (Kaiser-Permanete Plan) and fee-for service care, showing fewer physicians can handle more people in alternative organizational arrangements. Lewis [63] advocates different organizational delivery systems to eliminate excessive duplication of services.

Delivery system policy questions include: What if we changed the organization? What should it be changed to, and what can we predict about the effects? Will Health Maintenance Organizations (HMO's) improve the situation? To answer these questions we would need a model whose design does not depend on existing patterns of delivery, but allows representation of alternate systems.

Many models have been formulated around this problem, though none have attempted to represent the general design problem. We cite four such models:

Scheffler [97] formulated a non-linear mathematical programming model to search for an optimal size for a group practice. Schneider et. al. [100] investigated design of HMO's as mathematical programs in terms of three measures of effectiveness—(1) minimize cost to the subscriber, (2) minimize the number of professional manpower to serve a given set of subscribers (and maintain quality), and (3) maximize services provided by a given set of professionals; these require integer solutions. The research effort did obtain enough data to compute an illustrative output, and more analysis will be forthcoming.

Schuman, et. al. [103] developed an integer-mathematical programming problem for a hypothetical neighborhood health center and expressed the problem as one to determine the mix of personnel and services, the level of technology, and the amount of construction and community resources necessary to minimize the total "cost to society" of providing health services, subject to the constraints of quality of service, service requirements, available personnel, capacity of facilities, financial considerations, and relationships of personnel and services. Milly and Pocinki [69] developed a simulation model based on service-treatment stations; it will be used to evaluate alternative delivery system designs for neighborhood health centers, medical group practices and cooperative pre-paid group delivery systems.

None of the above models are able to incorporate health status measures in any explicit way. In the models, people arriving at services require resources. Their various diseases and disabilities are represented as demands for particular services. Only the Milly and Pocinki approach seems, at this time, capable of including other considerations such as patient satisfaction, and some connection with morbidity and mortality in the population served.

National Program Evaluation and Comparison

Two related policy problems are health program evaluation and health program comparison. Evaluation is a retrospective analysis of an existing health program to see if the effects produced by the input of resources justified, in some sense, the cost of those resources. Program comparison can apply either to a comparison of two (or more) program evaluations, to see if one program produced more effect than others for the same relative resource input; or to predict for several existing and/or proposed programs, which ones should produce the most effect per dollar, and thus help policy makers decide on priorities for program expansion or development when funds cannot support all programs. In addition, the design of a single new program may require comparisons of effects at different levels of funding, [18,39]. An important program evaluation effort was the HEW analysis of selected disease programs, which was designed to answer the policy question: "If additional money were to be allocated to disease control programs, which programs would show the highest payoff in terms of lives saved and disability prevented per dollar spent?" Grosse [39]. The resulting estimates and background analyses have influenced HEW allocations since 1966 [114]. For example, they helped in setting a high priority on educational activities to convince people to wear seat belts in motor vehicles.

The Indian Health Service constructed a "Q index" to determine priorities among several problem areas Miller [68]. The procedure was designed to put high priority on health conditions that tend to reduce total time lost due to the health problem, including lost time due to early death. The index tends to assign highest priority to accidents, because so many happen to children.

Levy [62] collected a set of operating data on the Illinois mental health program to determine if sufficient data could be acquired to evaluate the program. This led to the need for measuring the "quality of life" of discharged patients, but this relies on indirect evidence given by available statistics, i.e. the patient did not come back.

A complex program evaluation model, specific as to tuberculosis programs, was designed by Chorba and Sanders [15]. Benefits were calculated primarily from program savings resulting from prevention and early cure, so that expensive later treatment would not be needed.

As part of an ongoing Health Index Project (Bush, et al. [8,9]), a mathematical programming approach was developed by Chen and Bush [13] for the selection of health programs. This included budgetary constraints, specific programs and in effect selected people for program treatment so as to minimize the time people would spend in highly dysfunctional health status categories.

A mathematical programming model was also used to choose "best" actions in another specific health program, that of family planning, Correa and Beasley [17]. The measure of effectiveness was the difference between actual and desired children per family, and the formulation included six contraceptive methods and the costs of providing the contraceptives and doctor visits.

Correa [16], also demonstrated the use of mathematical programming in allocation of national resources toward either preventive care or curative treatment. Results suggested that all uncommitted resources should go into preventive services, to minimize death and disease.

III. COMPREHENSIVE HEALTH PLANNING MODELS

Regional Health Planning Decisions

The policy questions associated with regional (State and local) planning, include: maldistribution of health resources; unmet need for particular health services; specific problems such as infant mortality and air pollution; optimum utilization of present services; and state credentialling of manpower and certification of facilities and services.

There have been a few models designed specifically for regional planning use. Because of a lack of data, and lack of an organizational focus for the user of a model, such models are experimental at best. The models are used to speculate on how policy questions could be handled if proper data could be developed and if funds were available to convert the model design into operating reality. The designs are generally simulation models, rather than the econometric national policy models.

The discussion of models below is organized by main headings of model type, e.g. econometric, with related problems discussed in each section.

Adaptive Control Models

Adaptive control models which could encompass all health services in an area would logically seem to fit planning needs better than other model types; their application to health services has been discussed by Howland [49 and 50]. One of the most important features of adaptive control models is the concentration on information needed for decisions.

Figure 5 shows the conceptual elements of a control-oriented model. Assumptions include: (1) a comparison of input and output is made, to see what changes have occurred as a result of using the services; (2) the results of the comparison are used by a regulation function, which can then make adjustments to the system activities or system resources; and (3) the comparison of input and output is guided by objectives derived from goals. Thus, this representation assumes that the planning function can influence system performance by regulating the health system processes, or by changing the supply or resources.

Palmer, et al. [83] presents a planning system within which a model of the entire health services system would be useful. The planning system as a whole is considered to be an adaptive control system, with objectives serving as the basis for comparing actual data with desired results. Projects, persuasion, and recommendations for legislation become the control actions. The model would be used to predict the effects of the control actions, recognizing the long time lag before measurable effects can be expected.

Probability Models

The probability approach separates a population into different categories, and calculates how transitions will occur among the categories from one time "snapshot" to the next. This approach has been described by Navarro [75] and [77], Schach and Schach [96], Newheiser and Schoeman [78], and Nakamura [74]. To some extent, the models of disease development discussed by Bush, et al. [10], and Ortiz and Parker [80], and also the model of epidemic contagion by Thomas [107], could be considered in the regional planning context, even though the models do not include specific health resources. The next two paragraphs describe some of the approaches to defining the probabilities of transition.

General Health Status Categories. These models incorporate probabilities

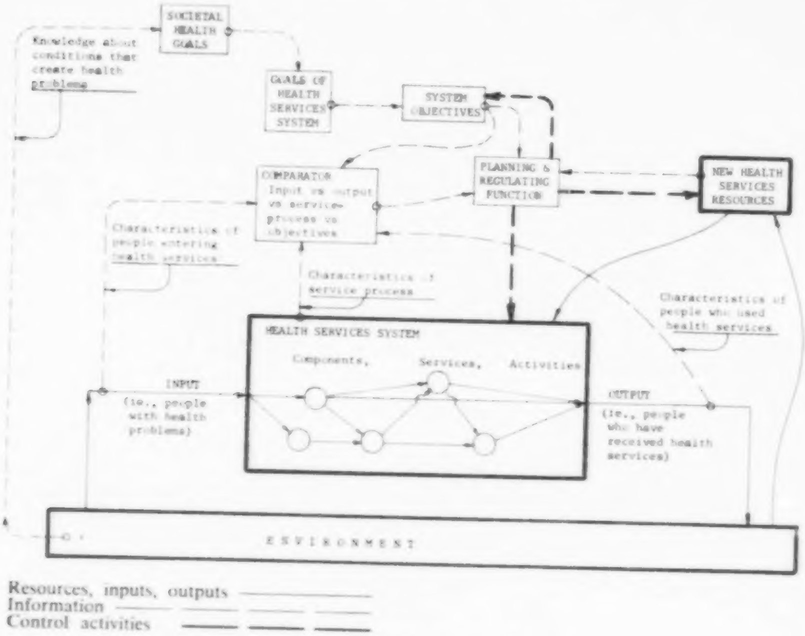

SOCIETAL HEALTH GOALS
Knowledge about conditions that create health problems
GOALS OF HEALTH SERVICES SYSTEM
SYSTEM OBJECTIVES
COMPARATOR Input vs output vs service-process vs objectives
PLANNING & REGULATING FUNCTION
NEW HEALTH SERVICES RESOURCES
Characteristics of people entering health services
Characteristics of service process
Characteristics of people who used health services
HEALTH SERVICES SYSTEM
Components, Services, Activities
INPUT (ie., people with health problems)
OUTPUT (ie., people who have received health services)
ENVIRONMENT
Resources, inputs, outputs
Information
Control activities

FIGURE 5—*Simplified Adaptive System*

of transition from one health state to another. Palmer [82], attempts to do this by postulating three sets of transition probabilities: one based on no health services being received, one based on "ideal" care being received, and a third based on actual services received, which is usually somewhere in between "no care" and "ideal care." Papell and Crocetti [84] assign people to "risk categories" which are close to general health status categories, but are primarily designed for an ambulatory care service.

Health Services Categories. This method of classification estimates the distribution of people receiving various health services at a given time. The probability estimates are for transitions to a different type of service between time periods. Navarro [75] used 6 medical care categories for a regional population: not under care, primary medical care, consultant medical care, hospital care, nursing home care, and domiciliary care. These "utilization" probabilities were used to estimate resource requirements for each type of care. The concept was extended to age-specific probabilities, decreased number of care-categories, and shortened time intervals, Navarro, et al. [77]. This formulation was simulated to obtain annual averages over a ten year period. Schach and Schach [96] used the same approach to estimate the averages and standard deviations of resources needed. The need in turn was based on the distribution of people using each type of health service.

These formulations still do not cover the full range of health services; mental, dental, and environmental are left out, as well as preventive and promotional activities. The distribution of services has not been represented either. The emphasis is basically an estimation of future resource requirements, assuming past utilization rates will continue to change in line with past trends. Data problems are almost insurmountable with probability models, even with aggregated categories.

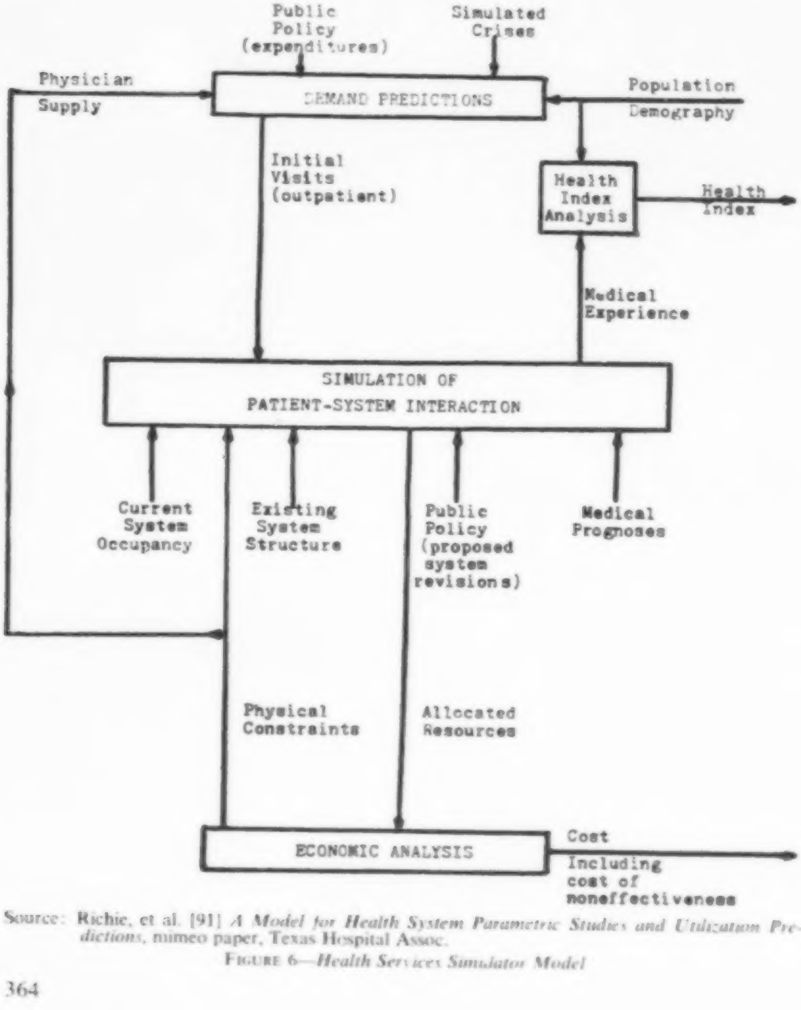
Utilization Simulation Models

Zemach [126] has designed a simulation model to combine the many factors related to production, utilization, and cost of health services. The model may be used to simulate a few proposed changes in the health system, if estimates on changes in utilization rates can be provided. By estimating future population, a forecast of resource requirements can be derived, assuming the utilization units per person and resource-service relationships remain about the same. The linkage to changes in health status as a result of resource useage is not included. Mental and environmental health services are not modeled.

The California State Office of Comprehensive Health Planning has developed a deterministic procedure for computing demand and supply estimates for health services (California [12]). A Personal Health Services module computes people who will be attended by health services, a Community Health Systems Analysis module computes the health system capacity, and an Evaluation of Community Health System module will compute various measures of system performance, including impact on health status. This model represents an ambitious attempt to relate data from many sources into an integrated framework. Much of the data is being derived from a 1964 study of a pre-paid group health plan in New York City, Avnet [4], and National Health Interview data. The model is very much disease-oriented and excludes dental, mental, and environmental services.

A model using a combination of model types has been developed by Richie, et al. [90] and [91], and a variation has been described by Lawless and Richie [60]. The overall procedure uses a regression analysis to predict entry of people into the health care system, by estimating initial visits for each of nine age groups. A simulation procedure then estimates useage of inpatient and outpatient care, to produce resource useage data. A series of equations then determines costs and compares resources needed against known resources available to identify shortages. The model is shown in a simplified flow chart in Figure 6. The procedure was originally designed to study effects of sudden "shocks" applied to the health services system, such as epidemics and disasters. The supply-demand estimates would be about in balance under normal system conditions (especially since the regression equations use past data). The major interest would lie in the supply shortages to be expected under sudden, unusually heavy demands on services. Work with the model has placed more emphasis on estimates under normal conditions, and suggestions have been made to use the model to gauge the impact of national health insurance, HMO's use of paramedical personnel, and other proposed changes.

A group at the University of British Columbia developed an aggregate simulation model of the health care system


Public Policy (expenditures)
Simulated Crises
Physician Supply
DEMAND PREDICTIONS
Population Demography
Initial Visits (outpatient)
Health Index Analysis
Health Index
Medical Experience
SIMULATION OF PATIENT-SYSTEM INTERACTION
Current System Occupancy
Existing System Structure
Public Policy (proposed system revisions)
Medical Prognoses
Physical Constraints
Allocated Resources
ECONOMIC ANALYSIS
Cost Including cost of noneffectiveness

Source: Richie, et al. [91] *A Model for Health System Parametric Studies and Utilization Predictions*, mimeo paper, Texas Hospital Assoc.

FIGURE 6—*Health Services Simulator Model*

in Vancouver, Canada (Milsum, et. al. [70]). While the Canadian health services system is very different from the U. S. system in many respects, the project is worthy of note because the health service model was built for regional planning purposes and is a part of a larger modeling effort designed to include other systems such as transportation, land-use, and education. In addition, the model includes a scale of social impact as a method for gauging psychological effects of not providing health services when people need (or want) them. The model simulates people in various disease categories, with a seriousness of illness rating scale constructed from 126 diseases. The general idea is that available resources are considerably short of demand, so a conscious effort is needed to allocate the available resources to high priority health problems, as derived from the seriousness scale. The social impact scale is then used to estimate the effect in terms of cases not treated. By building the health planning model as a part of a larger model, a simulation run could show the effect that changes in external environmental conditions have on health. By simulating for future time periods, the value of preventive services can be demonstrated.

The health portion of the overall Vancouver model is being adapted for the province of Quebec (Quebec [86]). Development efforts are now in progress to use data being generated as part of a new Quebec health care delivery system. Disease-specific data should eventually include virtually every transaction between individuals and health care providers. Initial model computations have been limited to insurance data which shows the use of physicians and hospitals by disease category. While the model does not include dental, mental, or environmental services, vital questions on hospital planning and delegation of duties to paramedical personnel are being simulated.

Econometric Models

An econometric model has been developed specifically for the use of Comprehensive Health Planning agencies at the areawide (sub-state) and statewide levels, Yett, et al. [120]. The model requires data for over 100 variables as well as some familiarity with econometric concepts and techniques. The general plan is to provide to Comprehensive Health Planning Agencies the model with a manual describing how to use it for forecasting and policy purposes. The model does not include dental, mental, or environmental health services, nor a measure of health status.

Another econometric approach to modeling a regional health system was developed by Davidson and Dahl [19] for the Minnesota Comprehensive State Health Planning Agency. Data from 1970 were obtained for 86 of Minnesota's 87 counties (the county containing the Mayo clinic was omitted to avoid distortion). Forecasts for 1975 and 1980 were computed under several assumptions. Major results were the estimates of requirements, by county, for health manpower, as well as estimates of patient days in hospitals, outpatient clinic visits, emergency visits, patient days in nursing homes, primary care physicians, and secondary care physicians.

Mathematical Programming Model

A model of a state formula for financing mental health services in New York was successful in analyzing related policy problems, Bodin, et. al. [7]. The analysts worked closely with state legislators to incorporate politically feasible value of county contributions in a shared-funding formula. The analysis produced an alternative proposal to one already in committee. The model was a mathematical program with a non-linear (quadratic) objective function. It used an optimizing criteria based on minimizing the deviations of projected future county expenditures from present expenditures, i.e. do not let the deviations vary too drastically or political feasibility will be impaired. While the close association between the analysts and the legislators produced a politically feasible analysis, the scope of the problem was extremely narrow. Alternatives for treating the mentally ill were not included in the analysis.

Ill-Health Generation Models

Two other models should be mentioned in connection with regional planning: one because of the attempt to relate environment conditions to changes in health, and the other because of the use of a different modeling technique.

The Environmental Systems Group at the University of California at Davis, Watt [116], has been developing a Model of Society, simulating the California air pollution environment. It computes expected medical effects on the population, as well as on plant growth, agriculture and meteorological conditions. The model does not include the health services system: it models the extreme lefthand portion of factors shown in Figure 1, the generation of ill-health conditions. In an associated study, Hickey [46] used a correlation analysis to associate air pollutants with chronic diseases and found strong evidence of such association.

Anderson [3] also studied the illness-generation factors and constructed a model of health services which included many demographic variables associated with poor health based on data from the 32 counties in the state of New Mexico. A form of statistical regression model, called path analysis, was used which depends on hypotheses concerning the direct and indirect effects of the variables on health delivery and on the health status of the population. Results showed that hospital beds per 100,000 population appears to have a small positive direct effect on the mortality index; per capita income appears to have a negligible effect on health; and unemployment has a large direct effect on the mortality due to accidents, on suicide, and on cirrhosis of the liver. Since the variables used are not considered controllable by the health services system, the results are difficult to use in giving guidance to health planners.

IV. COMMUNITY HEALTH SERVICE ALTERNATIVES

Community Health Service Decisions

Community health problems include: rat control, alcoholism, narcotics addiction; service delivery problems include: maternal and infant care, emergency medical care and ambulance services, coronary care, food sanitation, and blood bank inventory control. These problems require coordination of a variety of agencies in developing alternative solutions. The alternatives are sufficiently narrow so that modeling is quite feasible. Many health problems at the community level have rather well-defined treatment actions that can produce measurable effects. Once a responsible group in a community has either already started a program, or has determined that a health program action is needed, models can be used to choose among alternative actions, to monitor an on-going program, or to evaluate prior programs.

This section discusses briefly some of the models used in analyzing community health service program alternatives.

Blood Bank Models

Analysis of blood banking procedures are necessary due to the problems of shelf-life, cross-matching, and the implications of a blood shortage or errors in matching.

Rabinowitz and Valinsky [87] constructed a simulation model of an individual hospital blood bank, keeping track of 8 different blood types and incorporating the reservation policies actually in use. The model was used to test policies for managing the inventory. Frankfurter and Pegels [34] constructed a blood inventory forecast model for a hospital, based on past distributions of useage by day of week, and other variables. These models are used to aid blood bank managers in ordering and assigning units of blood within individual hospitals. Jennings [53, 54], Jelmert and Pegels [52], have modeled a distribution center responsible for allocating daily supplies to participating hospitals, and policies related to inter-hospital transfers of blood. Pegels and Jelmert [85] applied a probability analysis to calculate average wastage rates and average life of an entire city's inventory.

Emergency Medical Care Models

Emergency medical care systems have been modeled primarily by simulation. Typical decisions supported by models are the optimal number of ambulances to cover a service area (given probabilities of emergency incidents) and the optimal location of the ambulance dispatch centers so as to minimize response time to the scene of randomly occurring emergencies, New York City (Savas [95]), Detroit (Hall [42, 43]), and Los Angeles (Fitzsimmons [32, 33]). Stevenson [104] designed an analytical probability model to study the use of less efficient auxiliary vehicles when primary vehicles are all in use. A detailed simulation model of the total emergency medical service from time of injury to time of release from the emergency medical facility was developed by Hare and Wemple. [44].

Planning Alcoholism Programs

Holder and Hallan [47] describe the use of a simulation approach in planning alcoholism programs in 20 counties of North Carolina. The project identified ten intervention points for dealing with problem drinkers. Preliminary simulations were used to guide establishment of potentially effective programs. Performance was judged by changes in personal drinking behavior, as measured by data collected on clients served during 4 two-month periods.

Maternal and Infant Care Programs

Kennedy and Woodside [57] describe a maternal and infant care (MIC) simulation model. It was used to evaluate effects of changes in population or nursing staffs on project operations in three North Carolina counties. Four regression equations were developed to predict the number of visits a mother makes to the prenatal clinic and the number of infant illnesses. The link between needs and resources was based on the time required of all resources in the clinic to care for a patient. Initial data were obtained from an MIC clinic and the model was then tested against data from other clinics.

Family Planning Programs

O'Connor and Urban [79] reported on a model of patient flow in family planning programs in Atlanta, Georgia. The analysts worked with directors of family planning services to define specific flow patterns, data and costs in the model. By participating in actual building, program directors understood the model and the necessity for appropriate data. Projections obtained using the model aided in annual budget discussions.

Narcotics Control Models

A cost benefit analysis of various programs for treating heroin addicts was constructed by Maidlow and Berman [65]. The model evaluated two main programs: the therapeutic community as an abstinence approach, and methadone maintenance as a substitution approach. Direct costs, maximum benefits (if everyone were cured) and probabilistic net benefits (adjusting for drop-outs) were calculated for each treatment program. Most of the benefits were due to addicts refraining from stealing to support the habit.

A feedback model was used by Levin, et al. [61] to simulate the economics of narcotics control programs. This model searches for stable conditions among the supply, addict-population, and other variables associated with narcotics addiction problems. Results from the model suggested that a variety of community response actions would be best; i.e. police response without rehabilitation and education drives up prices and increases crime proportionately, rehabilitation alone leaves a tempting supply to reinfect the addicts; and community education alone is simply ineffective. Only if all three program responses are used does the model show some stability of conditions after a 3-year period.

V. INSTITUTIONAL PLANNING AND CONTROL MODELS

Facility Planning and Control Decisions

The planning decisions which are concerned primarily with an institution as a whole can be grouped into categories. The categories with references to models developed for these areas follows. As we will not describe these models, the reader is referred to [105] for an overall review.

1. New Facilities:
   - (a) Design decisions—how does architecture and layout of the physical plant affect hospital efficiency and patient care? [20, 22, 23, 92, 93, 111, 117]
   - (b) Location and size decisions—where in a given geographic area should new facilities be built? Should one large facility be built, or several smaller ones? [1, 24, 38, 71, 72, 99, 101]
2. Existing Facilities:
   - (a) Overall efficiency: management control decisions. [5, 21, 29, 30, 31, 66, 89, 109, 112]
   - (b) Patient admission decisions—how should elective admissions be scheduled, when combined with stochastic elements of emergency admissions? [108]
   - (c) New services decisions—how would addition of a new intensive care unit, for example, affect operations of a hospital as a whole? [25, 41]
   - (d) Staff assignments and bed-use decisions—how to allocate staff and beds to improve service and use skills most effectively? [115, 119]
   - (e) Department/Clinic workload, staffing, inventory, and scheduling decisions. [2, 45, 64, 88, 98, 106]

VI. ASSESSMENT OF HEALTH SYSTEM MODELS FOR POLICY ANALYSIS

Model Types Versus Level of Decision

The preceding sections show that many health models have been developed at each decision level. No one model type has been determined to be best for any given decision level. Rather, researchers with a particular way of viewing the health care system have applied familiar methodology to their areas of interest. For example, staff of the Johns Hopkins University have been the source for many probability models of institutional problems, with extensions to community services and then to areawide planning. Work from the Research Triangle Institute has applied simulation models to institutional and community services, with extension to national health manpower decisions. A group at the University of Southern California has specialized in econometric models of the nation's health services and manpower policies, with extensions to regional planning, while researchers at the Massachusetts Institute of Technology are working on feedback models of large scale systems, with applications to health care systems. Many individuals, as noted in the previous sections, have also pioneered and applied specific models to the health area.

Our general assessment of models in the health services field is, however, that most of the modeling efforts are used infrequently or not at all by the appropriate decision makers. Although no specific answer can explain the gap between model developer and policy maker, we offer the following comments.

The insights and understanding to be gained in the building of a health model are valuable in themselves. Researchers have constructed econometric models partly, if not mostly, to gain an understanding of economic market forces that have been at work to produce the current pattern of physicians, nurses, hospitals, prices, and government expenditures. Others have constructed detailed simulation models of hospitals and nursing homes at least partly to understand how the work tasks of employees, such as nurses, affect the overall profits and output of the institution. This may help explain why some modeling efforts do not get beyond the design stage: the modeler already gained enough of the understanding he was after, and foresaw too many data problems to justify further work.

There are serious measurement limitations in health models as there is no agreed upon measure of health status that can summarize physical, dental and mental health conditions. An accepted overall measure of health would help modeling be more useful in policy decisions.

The objectives of most of the models have been specified by the modelers, rather than by policy makers. The policy makers must be involved in the analysis if they want useful results. In health this means decision makers from HEW to County Health Departments. Also, in the health field (as well as in other social fields) data are not collected in consistent categories suitable for model use. There is no data at all for many variables and parameters that must be specified. Models tend to build-in the present system and are therefore difficult to use in predicting effects of major changes in that system; some policy questions arise because there is a suspicion that past relationships should change. Proposals, such as national health insurance and HMO's are advanced concepts which should, if successful, substantially alter the entire health services system. Furthermore, as the Swedish economist Myrdal [73] points out, parameter estimation depends on the time period chosen, the population included and the geographic area encompassed. Parameter estimates are not constant over time or area. Therefore, a prediction of effects using model parameters derived from past relationships can be inaccurate at best and may be highly misleading.

Prognosis and Conclusion

Are any limitations so serious that a person in a position of real responsibility should not even consider using a model and its results as major guides to health legislation, planning or operating decisions? There are many current efforts which will substantially reduce some of the limitations. For example, research on health status measures will probably produce a useful, generally accepted measure of health condition. The Federal-State-Local Cooperative Health Statistics program will eventually standardize data definitions and categories, and produce data useable in models. Problems of inadequate objectives and semantics can and must be overcome by establishing closer liaison between modelers and policy makers.

It seems that health system modeling will continue and even expand, mostly as part of academic studies. The utility of health models as policy analysis instruments will depend on developing a rapport between analyst and persons in responsibility, probably requiring new organizational arrangements. The flavor of what will be needed has been suggested by Dr. Charles D. Flagle, as given in Horvitz, et al. [48]:

"Our pediatric clinics simulation model is one in which the resident who runs the clinic sits at the computer terminal and decides experimentally whether he wants four doctors or five in a simulated morning session, or decides whether he would like to have only a walk-in clinic instead of walk-in and scheduled clinic or decides to see what would happen if some day twice as many people come in. The simulation model demonstrates the effects of his decisions in terms of delays and congestion.

"The point is from the time the physician administrator sits down at that terminal, the model is his and not ours. It was ours in the beginning, and when it was only ours he didn't care very much. But having made it a part of his own activity and working on it in the context of the very organization in which he has the authority to make decisions it becomes very much a part of his existence. We might ask, How do we reproduce that situation in the context of regional planning or comprehensive health planning? How do we set up a health planning game, as it is possible in cases of military war gaming or industrial management games?

"We need situations with a scenario, the running dialogue. We need to have analysts and decision-makers in the same room with all sorts of displays on the wall, displays coming quickly from the computer, showing people the results of the decisions they have made in their role in the simulation. Such an alliance may permit us to avoid some of the connotations of the word 'utopian,' the notion of naivete, of dreaminess, of leaving out something important. Perhaps we can avoid this if we can bring into effect the necessary coalition of practical and theoretical decision-maker and scientist."

References

1. Abernathy, W. J., and Hershey, J. C., "A Spatial-Allocation Model for Regional Health-Services Planning," *Operations Research*, Vol. 20, No. 3, May-June 1972; pps. 629–642.
2. Abranovic, W. A., and Wallace, W. A., "A Computer Simulation Approach for Planning in Hospital Ancillary Services," *Socio-Economic Planning Sciences*, Vol. 5, No. 5, Oct., 1971; pps. 429–448.
3. Anderson J. G., "Causal Model of a Health Services System," *Health Services Research*, Vol. 7, No. 1, Spring, 1972; pps. 23–42.
4. Avnet, H. H., *Physician Service Patterns and Illness Rates*, Group Health Insurance, Inc., 1967.
5. Baligh, H. H., and Laughhunn, D. J., "An Economic and Linear Model of the Hospital," *Health Services Research*, Vol. 4, No. 4, Winter, 1969; pps. 293–303.
6. Blum, H. L., "Priority Setting for Problems, Solutions, and Projects, by Means of Selected Criteria," *International Journal of Health Services*, Vol. 2, No. 1, 1972.
7. Bodin, L. D., Carroll, T. O., Lee, A., and Stout, S., "Financing Mental-Health Services in the State of New York," *Operations Research*, Vol. 20, No. 5, Sept.-Oct., 1972; pps. 942–954.
8. Bush, J. W., Chen, M., and Patrick, D., *Health Index Project*, Mimeo paper of Department of Community Medicine, University of California, San Diego, California; 1972.
9. Bush, J. W., Chen, M. M., and Patrick, D. L., "Social Indicators for Health, Based on Function Status and Prognosis," *1972 Proceedings of Social Statistics Section, American Statistical Association*; Aug. 14–17, 1972, Montreal, Canada; pps. 71–80.
10. Bush, J. W., Chen, M. M., and Zaremba, J., "Estimating Health Program Outcomes Using a Markov Equilibrium Analysis of Disease Development," *American Journal of Public Health*, Vol. 61, No. 12, Dec., 1971; pps. 2362–2375.
11. Bush, J. W., Fanshel, S., and Chen, M. M., "Analysis of a Turberculin Testing Program Using a Health Status Index," *Socio-Economic Planning Sciences*, Feb., 1972, Vol. 6, No. 1; pps. 49–68.
12. California State Office of Comprehensive Health Planning, *Community Health System Analysis and Evaluation Model*, CHP Program Analysis and Evaluation Bureau (mimeo), Sept., 1972.
13. Chen, M. M., and Bush, J. W., *A Mathematical Programming Approach for Selecting an Optimum Health Program Case Mix*, Paper presented before meeting ORSA, Anaheim, California, Oct. 27, 1971.
14. Chorba, R. W., Miller, J. E., and Sanders, J. L., *Models for the Description and Evaluation of Acute and Chronic Disease Programs*. Mimeo paper, University of Arizona, 1969.
15. Chorba, R. W., and Sanders, J. L., "Planning Models for Tuberculosis Control Programs," *Health Services Research*, Vol. 6, No. 2, Summer, 1971; pps. 144–164.
16. Correa, Hector, "Health Planning," *KYKLOS*, Vol. 20, 1967; pps. 909–923.
17. Correa, H., and Beasley, J. D., "Mathematical Models for Decision-Making in Population and Family Planning," *American Journal of Public Health*, Vol. 61, No. 1, Jan., 1971; pps. 138–151.
18. Crystal, R. A., and Brewster, A. W., "Cost Benefit and Cost Effectiveness Analyses in the Health Field: An Introduction," *Inquiry*, Vol. 3, No. 4, Dec., 1966, Chicago; pps. 3–13.
19. Davidson, G., and Dahl, T., *A Macro-Statistical Model of the Health Sector for the State of Minnesota*, Minnesota Systems Research, Inc., Minn., Minn., Feb., 1973, Vol. I, Introduction and Summary, Vol. II, Technical Report.
20. Delon, G. L., "A Methodology for Total Hospital Design," *Health Services Research*, Vol. 5, No. 3, Fall, 1970; pps. 210–223.

21. Delon, G. L., *Long Range Planning for Mental Health Institutions Using Feedback Dynamics*, Mimeo paper, Health Systems Research Group, Georgia Institute of Technology, prepared for 38th meeting, ORSA, Oct., 1970, Detroit, Michigan.

22. Delon, G. L., and Smalley, H. E., "A Stochastic Model for Inter-departmental Traffic," *Health Services Research*, Vol. 5, No. 3, Fall, 1970, pps. 196–209.

23. Delon, G. L., et al., *Quantitative Methods for Evaluating Hospital Designs*, Final Research Report: National Center for Health Services Research & Development, April, 1970, Avail. from NTIS, PB 190389.

24. DeVise, P., "Viewpoints: Methods and Concepts of an Interdisciplinary Regional Hospital Study," *Health Services Research*, Vol. 3, No. 3, Fall, 1968; pps. 166–173.

25. Emlet, H. E., *Impact of Reduced Physician Time per Patient on Cost-Benefits of Multi-phasic Health Screening*, Mimeo paper, Analytic Services, Inc., presented to 37th Meeting, ORSA, April 20–22, 1970, Washington, D.C.

26. Fanshel, S., and Bush, J. W., "A Health Status Index and Its Application to Health Services Outcomes," *Operations Research*, Vol. 18, No. 6, Nov.-Dec., 1970; pps. 1021–1066.

27. Feldstein, M. S., "An Econometric Model of the Medicare System," *Quarterly Journal of Economics*, Vol. 85, No. 1, Feb., 1971; pps. 1–20.

28. Feldstein, P. J., and Kelman, S., "A Framework for an Econometric Model of the Medical Care Section," *Empirical Studies in Health Economics*, Klarman, H. E., (ed.) Baltimore, Johns Hopkins Press, 1970; pps. 171–190.

29. Fetter, Robert and Thompson, John, "A Decision Model for the Design and Operation of a Progressive Patient Care Hospital," *Medical Care*, Vol. VII, No. 6, Nov.-Dec., 1969; pps. 450–462.

30. Fetter, R. B., and Thompson, J. D., *A Planning Model for the Delivery of Health Services*, Mimeo paper, Department of Administrative Services, Yale University, 1971.

31. Fetter, R. B., "*Planning Models for Health Care Systems*," Mimeo paper, Technical Report #54 of Health Services Research Program, Yale University, 1972.

32. Fitzsimmons, J. A., "A Methodology for Emergency Ambulance Deployment," *Management Science*, Vol. 19, No. 6, Feb., 1973; pps. 627-636.

33. Fitzsimmons, J. A., "An Emergency Medical System Simulation Model," *Proceedings of 1971 Winter Simulation Conference*, Sponsored by ACM/AIIE/IEEE/SHARE/SC:/TIMS.

34. Frankfurter, G. M., and Pegels, C. C., *Short Term Blood Inventory Forecast Model*, Mimeo paper, School of Management, State University of New York at Buffalo; 1970.

35. Fromm, E., and Goodman, R. M., editors: *Record of the Symposium on Future Patterns in Health Care Delivery*, Held May 5, 1972, Philadelphia, Pa.; sponsored by IEEE, Philadelphia Section.

36. Garfield, S. R., "The Delivery of Medical Care," *Scientific American*, Vol. 222, No. 4, April, 1970; pps. 15–23.

37. Glazer, Nathan, "Paradoxes of Health Care," *The Public Interest*, No. 22, Winter, 1971; pps. 62–77.

38. Gross, P. F., *Urban Health Disorders, Spatial Analysis and the Economics of Health Facility Location*, Mimeo paper of University of Saskatchewan, Canada; presented at 40th meeting, ORSA, Oct. 28, 1971, Anaheim, California.

39. Grosse, R. N., "Problems in Resource Allocation in Health," Chapter 22 of *Public Expenditures and Policy Analysis*, edited by Haveman & Margolis, Markham Publ. Co., Chicago, 1970; pps. 518–548.

40. Grossman, M., *The Demand for Health: A Theoretical and Empirical Investigation*, National Bureau of Economic Research, Columbia University Press, New York, 1972.

41. Gurfield, R. M., and Clayton, J. C., *Analytic Hospital Planning: A Pilot Study of Resource Allocation Using Mathematical Programming in a Cardiac Unit*, RAND memorandum RM-5893-RC, April, 1969; Santa Monica, California.

42. Hall, W. K., "Management Science Approaches to the Determination of Urban Ambulance Requirements," *Socio-Economic Planning Sciences*, Vol. 5, No. 5, Oct., 1971; pps. 491–499.

43. Hall, W. K., "The Application of Multi-function Stochastic Service Systems in Allocating Ambulances to an Urban Area," *Operations Research*, Vol. 20, No. 3, May-June, 1972; pps. 558–594.

44. Hare, R. R., and Wemple, A. B., *Model for the Analysis of Emergency Medical Service Systems*, Technical Report 555, Operations Research, Inc., June, 1970.

45. Haussman, R. K. D., "Waiting Time as an Index of Quality of Nursing Care," *Health Services Research*, Vol. 5, No. 2, Summer, 1970; pps. 92–105.

46. Hickey, R. J., Boyce, D. E., Harner, E., and Clelland, R. C., *Ecological Statistical Studies on Environmental Pollution and Chronic Disease in Metropolitan Areas of the U.S.*, Regional Science Research Institute Discussion paper No. 35, Oct., 1970; Philadelphia, Pa.

47. Holder, H. D., and Hallan, J., "Systems Approach to Planning Alcoholism Programs in North Carolina," *American Journal of Public Health*, Vol. 62, No. 10, Oct., 1972; pps. 1415–1421.

48. Horvitz, D. G., Kennedy, F. D., Shah, B. V., and Jones, D. C., "Methods for Building a Health Manpower Simulation Model," in *Proceedings and Report on Conference on a Health Manpower Simulation Model*, Vol. 2; Bureau of Health Manpower Education; Dec., 1970.

49. Howland, Daniel, "A Regulatory Model for System Design and Operation," *Operational Research and the Social Sciences*, Lawrence, J. R., editor; Travistock Pub., London, 1966; Chapter 36, pps. 505–514.

50. Howland, Daniel, "Toward a Community Health System Model," *Systems and Medical Care*, Chapter 11; edited by Sheldon, Baker, and McLaughlin, MIT Press, Cambridge, Mass., 1970; pps. 268–289.

51. Jaeger, B. J., "Evaluating National Hospital Policy: A Systems Approach," *Health Services Research*, Vol. 7, No. 1, Spring, 1972; pps. 11–22.

52. Jelmert, A. E., and Pegels, C. C., *A Heuristic Distribution and Allocation Model for Whole Blood*, Mimeo paper, School of Management, State University of New York at Buffalo; June, 1969.

53. Jennings, J. B., "Blood Bank Inventory Control," *Analysis of Public Systems*, Drake, Keeney, & Morse, editors; MIT Press, Cambridge, Mass., 1972; Chapter 11, pps. 216–234.

54. Jennings, J. B., "Blood Bank Inventory Control," *Management Science*, Vol. 19, No. 6, Feb., 1973; pps. 637–645.

55. Joyce, Jon M., "More on Health Care Paradoxes," *Policy Sciences*, Vol. 3, No. 1, March, 1972; pps. 107–115.

56. Kane, J., Thompson, W., and Vertinsky, I., "Health Care Delivery: A Policy Simulator," *Socio-Economic Planning Sciences*, Vol. 6, No. 3, June, 1972; pps. 283–293.

57. Kennedy, F. D., and Woodside, M. B., "Vol. III: The Maternal and Infant Care Simulation Model," Final Report: *The Development of a Simulation Model of a Community Health Service System*, Research Triangle Institute, North Carolina, to HEW, Division of Community Health Services, March, 1968.

58. Kissick, W. L., editor, "Dimensions and Determinants of Health Policy," *Milbank Memorial Fund Quarterly*, Vol. XLVI, No. 1, Jan., 1968, Part 2. Proceedings of Seminar, Jan.-June, 1966.

59. Klarman, H. E., "Economic Aspects of Projecting Requirements for Health Manpower," *The Journal of Human Resources*, Vol. IV, No. 3, Summer, 1969, pps. 360–376.

60. Lawless, R. W., Williams, L. H., and Richie, C. G., "A Sensitivity Analysis Tool for Simulation with Application to Disaster Planning," *Simulation*, Vol. 17, No. 6, Dec., 1971, pps. 217–223.

61. Levin, G., Hirsch, G., and Roberts, E., "Narcotics and the Community: A System Simulation," *American Journal of Public Health*, Vol. 62, No. 6, June, 1972; pps. 861–873.

62. Levy, L., "An Evaluation of a Mental Health Program by Use of Selected Operating Statistics," *American Journal of Public Health*, Vol. 61, No. 10, Oct., 1971; pps. 2038–2045.

63. Lewis, I. J., "Government Investment in Health Care," *Scientific American*, Vol. 224, No. 4, April, 1971; pps. 17–25.

64. Liebman, J. S., Young, J. P., and Bellmore, M., "Allocation of Nursing Personnel in an Extended Care Facility," *Health Services Research*, Vol. 7, No. 3, Fall, 1972; pps. 209–220.

65. Maidlow, S. T., and Berman, H., "The Economics of Heroin Treatment," *American Journal of Public Health*, Vol. 62, No. 10, Oct., 1972; pps. 1397–1406.

66. McKnight, E. M., and Steorts, R. C., *Nursing Home Simulation Model, Final Report*, National Center for Health Services Research and Development, June, 1969; Available from NTIS, PB190795.

67. Meadows, D. H., Meadows, D. L., Randers, J., and Behrens, W. W., *The Limits to Growth*, Universe Books, New York, 1972.

68. Miller, J. E., "An Indicator to Aid Management in Assigning Program Priorities," *Public Health Reports*, Vol. 85, No. 8, Aug., 1970; pps. 725–730.

69. Milly, G. H., and Pocinki, L. S., *A Computer Simulation Model for Evaluation of the Health Care Delivery System*, Final Report, June, 1970, Available from NTIS, Springfield, Va., PB195132.

70. Milsum, J., Uyeno, D., Vertinski, I., and Will, H., "Vancouver Regional Health Planning Model," *Proceedings: 1971 Winter Simulation Conference on Applications of Simulation*, held Dec. 8–10, 1971, New York City; Sponsored by: ACM/AIIE/IEEE/SHARE/SCi/TIMS.

71. Morrill, R. L., and Earickson, R., "Locational Efficiency of Chicago Hospitals: An Experimental Model," *Health Services Research*, Vol. 4, No. 2, Summer, 1969; pps. 129–141.

72. Morrill, R. L., and Earickson, R., "Variation in the Character and Use of Chicago Area Hospitals," *Health Services Research*, Vol. 3, No. 3, Fall, 1968; pps. 224–238.

73. Myrdal, Gunnar, "How Scientific are the Social Sciences?" *Bulletin of the Atomic Scientists*, Vol. 29, No. 1, Jan., 1973; pps. 31–37.

74. Nakamura, M., *On a Markovian Model of Health Services Systems*, Mimeo paper, University of Alberta; presented at ORSA/TIMS/AIIE meeting, Nov. 8–10, 1972.

75. Navarro, V., "A Systems Approach to Health Planning," *Health Services Research*, Vol. 4, No. 2, Summer, 1969; pps. 96–111.

76. Navarro, V., and Joroff, S., *Medical Manpower: A Multivariate Analysis of the Distribution of Physicians in Urban United States*, Mimeo paper,

Department of Medical Care and Hospitals, Jons Hopkins University, 1970.

77. Navarro, V., Parker, R., and White, K. L., "A Stochastic and Deterministic Model of Medical Care Utilization," *Health Services Research*, Vol. 5, No. 4, Winter, 1970; pps. 257–342.

78. Newheiser, J. R., and Schoeman, M. E. F., "A Stochastic Model for Health Care Resource Planning," *Socio-Economic Planning Sciences*, Vol. 6, No. 2, April, 1972; pps. 197–213.

79. O'Connor, R. W., and Urban, G. L., "Using a Model as a Practical Management Tool for Family Planning Programs," *Amer. Journal of Public Health*, Vol. 62, No. 11, Nov., 1972; pps. 1493–1500.

80. Ortiz, J., and Parker, R., "A Birth-Life-Death Model for Planning and Evaluation of Health Services Programs," *Health Services Research*, Vol. 6, No. 2, Summer, 1971; pps. 120–143.

81. Packer, A. H., (Research Triangle Institute of North Carolina) "Applying Cost-Effectiveness Concepts to the Community Health System," *Operations Research*, Vol. 16, 1968; pps. 227–253.

82. Palmer, B. Z., and Goldman, C. I., *Design of a Simulation Model of Personal Health Services in an Urban Area*, Final Report to New Jersey Dept. of Health, Government Studies & Systems, Inc., March, 1970.

83. Palmer, B. Z., Sisson, R. L., Kyle, C., and Hebb, A., *An Advanced Health Planning System*, report by Government Studies & Systems, Inc. to New Jersey Dept. of Health; distributed by National Technical Information Services, U.S. Dept. of Commerce, Va., No. PB210 647.

84. Panell, B. M., and Crocetti, A. F., *CINCH: A Model of Practice for Community Health*, Mimeo paper, Dept. of Community Medicine, Mount Sinai School of Medicine, City University of New York, Aug., 1972.

85. Pegels, C. C., and Jelmert, A. E., "An Evaluation of Blood-Inventory Policies: A Markov Chain Application," *Operations Research*, Vol. 18, No. 6, Nov.-Dec., 1970; pps. 1087–1098.

86. Quebec Provincial Government, *MEDICS: A Program to Model the Health Care Delivery System of Quebec*, Ministere Des Affaires Sociales, Direction Generale De La Planification; September, 1972.

87. Rabinowitz, M., and Valinsky, D., "Simulation of Hospital Blood Banking," *Proceedings of 7th ACM Urban Symposium, on Computers and Urban Society*, Oct. 27, 1972, New York City.

88. Rath, G. L., Balbas, J. M. A., Ikeda, T., and Kennedy, O. G., "Simulation of a Hematology Department," *Health Services Research*, Vol. 5, No. 1, Spring, 1970; pps. 25–35.

89. Rath, G. J., Kennedy, O. G., and Lapp, R. E., *Developing a Composite Hospital Simulation Model*, Design and Development Center Report DDC–69–8–1, Northwestern University, Evanston, Ill., August, 1969.

90. Richie, C. G., Dove, H. G., Moore, W. S., and Edwards, S. A., "Developing a Health Services Simulator," *Proceedings of the 1971 Winter Simulation Conference on Applications of Simulation*, held Dec. 8–10, 1971, New York City; pps. 434–446. Sponsored by ACM AIIE IEEE/SHARE SCi TIMS.

91. Richie, C. G., Lawless, R. W., Moore, W. S., and McMillen, R. H., "*A Model for Health System Parametric Studies and Utilization Predictions*," Mimeo paper, Health Resources Planning Unit, Texas Hospital Association, San Antonio, Texas; presented at 38th meeting, ORSA, Oct. 30, 1970, Detroit, Michigan.

92. Rikkers, R. F., "Effect of Spoke Design on Occupancy: A Simulation Model," *Health Services Research*, Vol. 5, No. 3, Fall, 1970; pps. 233–247.

93. Rosenbloom, Arnold, "Systems Applications to Health Care Delivery Systems," *The Recognition of Systems in Health Services*, Chacko, editor; Operations Research Society of America, Arlington, Va., 1969; pps. 275–289.

94. Rosett, R. N., and Huang, L., *Analysis of Demand for Hospital and Physician Services*, Paper presented to joint session of Health Economics Research Organization and the American Economic Association, New Orleans, La., Dec. 28, 1971.

95. Savas, E. S., "Simulation and Cost-Effectiveness Analysis of New York's Emergency Ambulance Service," *Management Science*, Vol. 15, No. 12, Aug., 1969; pps. B–608 to B–627.

96. Schach, E., and Schach, S., "A Continuous Time Stochastic Model for the Utilization of Health Services," *Socio-Economic Planning Sciences*, Vol. 6, No. 3, June, 1972; pps. 263–272.

97. Scheffler, R. M., *Proposed Models for the Optimal Design of Health Care Delivery Systems*, (mimeo) Paper presented at ORSA meeting, Anaheim, California, Oct. 27, 1971.

98. Schmitz, H. H., and Kwak, N. K., "Monte Carlo Simulation of Operating Room and Recovery Room Usage," *Operations Research*, Vol. 20, No. 6, Nov.-Dec. 1972; pps. 1171–1180.

99. Schneider, J. B., and Symons, J. G., *Regional Health Facility System Planning: An Access Opportunity Approach*, Regional Science Research Institute, Discussion Paper No. 48, July, 1971; Philadelphia, Pa.

100. Schneider, D., Roberts, S., and Kilpatrick, K., *A Systems Analysis of Health Maintenance Organizations*, Mimeo paper, Fall, 1972, Paper delivered at Joint Meeting of Operations Research Society, TIMS, AIIE, Atlantic City, New Jersey, Nov. 10, 1972.

101. Schultz, G. P., "The Logic of Health Care Facility Planning," *Socio-Eco-*

nomic Planning Sciences, Vol. 4, No. 3, Sept., 1970; pps. 383–393.

102. Schwartz, William, B., "Policy Analysis and the Health Care System," *Science*, Vol. 177, Sept. 15, 1972; pps. 967–969.

103. Shuman, L. J., Young, J. P., and Naddor, E., "Manpower Mix for Health Services: A Prescriptive Regional Planning Model," *Health Services Research*, Vol. 6, No. 2, Summer, 1971; pps. 103–119.

104. Stevenson, K. A., "Emergency Ambulance Transportation," *Analysis of Public Systems*, Drake, Keeney, and Morse, editors; MIT Press, Cambridge, Mass., 1972; Chapter 8, pps. 133–150.

105. Stimson, D. H., and Stimson, R. H., *Operations Research in Hospitals: diagnosis and prognosis*, Hospital Research and Educational Trust, Chicago, Ill., 1972.

106. Sumner, A. T., and Hsieh, R. K. C., "Long-Range Prediction of Examining Room Requirements," *Health Services Research*, Vol. 7, No. 3, Fall, 1972; pps. 221–230.

107. Thomas, M. E., *A Stochastic Model for the Analysis of Epidemics*, Mimeo paper, University of Florida Industrial and Systems Engineering Dept., presented at 35th meeting, ORSA, June 17–20, 1969, Denver, Colorado.

108. Thompson, J. D., and Fetter, R. B., "Economics of Occupancy with Varying Mixes of Private and Other Patient Accommodations: A Simulation," *Health Services Research*, Vol. 4, No. 1, Spring, 1969; pps. 42–52.

109. Thompson, J. D., and Fetter, R. B., "The Application of Simulation to Hospital Planning," *Proceedings on Simulation in Business and Public Health*, sponsored by American Statistical Assoc. and Public Health Assoc., New York Area Chapters; 1966; pps. 205–210.

110. Torrance, G. W., Thomas, W. H., and Sackett, D. L., "A Utility Maximization Model for Evaluations of Health Care Programs," *Health Services Research*, Vol. 7, No. 2, Summer, 1972, pps. 118–133.

111. Tung, A., Parti, E. W., and Wong, A. K. C., *Hospital System Design Simulator*, Final research report, National Center for Health Services Research & Development, July, 1971. Available from NTIS PB201 132.

112. Turner, J. B., Moser, B. A., Colley, J. L., and Collins, R. L., *Computer Simulation Models of Nursing Homes*, Final Report for Public Health Service, Aug., 1967. Available from NTIS PB190 794.

113. U.S. Department of Health, Education, and Welfare, *Towards a Comprehensive Health Policy for the 1970's—A White Paper*, May, 1971.

114. U.S. Department of Health, Education, and Welfare, Office of Assistant Secretary for Program Coordination, *Program Analysis: Selected Disease Control Programs*, Sept., 1966.

115. Warner, D. M., and Prawda, J., "A Mathematical Programming Model for Scheduling Nursing Personnel in a Hospital," *Management Science*, Vol. 19, No. 4, Dec., 1972, Part 1; pps. 411–422.

116. Watt, K., (principal investigator), "A Model of Society," *Simulation*, Vol. 14, No. 4, April, 1970; pps. 153–164.

117. Weeks, J., and Best, G., "Design Strategies for Flexible Health Sciences Facilities," *Health Services Research*, Vol. 5, No. 3, Fall, 1970; pps. 263–284.

118. Wirick, G. C., "A Multiple Equation Model of Demand for Health Care," *Health Services Research*, Vol. 1, No. 4, Winter, 1966; pps. 301–346.

119. Wolfe, H. B., Iskander, M., and Raffin, I., "A Study of Obstetrical Facilities," *The Recognition of Systems in Health Services*, Chacko, editor, ORSA publication, 1969; pps. 367–392.

120. Yett, D. E., Drabek, L., Intriligator, M. D., and Kimbell, L. J., "A Macroeconometric Model for Regional Health Planning," *Economic and Business Bulletin*, Fall, 1971, Vol. 24, No. 1; Temple University; pps. 1–21.

121. Yett, D. E., Drabek, L., Intriligator, M., and Kimbell, L., "Conceptualization of a Health Manpower Simulation Model," in *Proceedings and Report of Conference on a Health Manpower Simulation Model*, Vol. 1, Bureau of Health Manpower Education, Dec., 1970.

122. Yett, D. E., Drabek, L., Intriligator, M., and Kimbell, L., "Health Manpower Planning: An Econometric Approach," *Health Services Research*, Vol. 7, No. 2, Summer, 1972; pps. 134–147.

123. Yett, D. E., Drabek, L., Intriligator, M. D., and Kimbell, L. J., *The Preliminary Operational HRRC Microsimulation Model*, Final Report to Division of Health Manpower Intelligence, Human Resources Research Center, USC, Los Angeles, California, 1973.

124. Yett, D. E., Drabek, L., Intriligator, M., and Kimbell, L., *The Use of an Econometric Model to Analyze Selected Features of National Health Insurance Plans*, Paper presented at a joint session of the Health Economics Research Organization and the American Economic Association, New Orleans, Louisiana, Dec. 28, 1971.

125. Yett, D. E., Drabek, L., Intriligator, M., and May, J., *Determination of Economic Attributes of Community Health Profiles and Health Indices*, May, 1970, mimeo. Available as Technical Paper No. 8, Community Profile Data Center, Community Health Service, HSMHA.

126. Zemach, R., "A Model of Health-Service Utilization and Resource Allocation," *Operations Research*, Vol. 18, No. 6, Nov.-Dec., 1970.

Chapter 12

Econometric Models for Policy Analysis

By

E. Philip Howrey

SUMMARY 377

I. INTRODUCTION 377

- What is Econometrics? 377
- Uses of Econometric Models 382
- Types of Econometric Models 384

II. NATIONAL ECONOMETRIC MODELS 386

- An Expository National Econometric Model 386
- Uses of a National Econometric Model 388
- Operational National Econometric Models 389

III. REGIONAL ECONOMETRIC MODELS 390

- An Expository Regional Econometric Model 391
- Uses Regional Econometric Models 393
- Operational Regional Econometric Models 394

IV. FORECASTING AND POLICY ANALYSIS: AN EXAMPLE 397

- Forecasting With an Econometric Model 398
- Policy Analysis With an Econometric Model 401
- Concluding Remarks 402

REFERENCES 403

- Econometric Methods 403
- National Econometric Models 404
- Regional Econometric Models 405
- Miscellaneous 406

APPENDIX: Construction and Use of Econometric Models 406

- Econometric Models 406
- Forecasting, Analysis and Control 412

Econometric Models for Policy Analysis

SUMMARY

The purpose of this chapter is to introduce the policymaker to econometric modeling and to indicate how econometric models can be used in forecasting and policy analysis. It is readily apparent that this chapter is substantially different from the other chapters of this *Guide* in that econometric modeling is a set of analytical techniques rather than a subject area such as urban economics or waste management. Econometric methods have been used to develop models in virtually all of the problem areas which are studied in the *Guide*.

Two areas in which econometric models have had or are expected to have an important impact need to be covered explicitly. These are national economic models and regional economic models. The purpose of these models is typically to forecast the level of economic activity as measured by gross national product for example and to provide a framework for the analysis of alternative tax and expenditure policies.

The discussion of this chapter is framed primarily in terms of what model builders generally hope to accomplish through their modeling efforts. Ideally, this would be accompanied by examples of the ways in which econometric models have been used in the actual formulation of economic policy. Unfortunately, it is virtually impossible to document the role of econometric models in policy design. The reason for this is that the use of models in the policy-making process is seldom formalized. This is not to say that models are not used; rather they are used in a rather informal way.

In order to provide the policy-maker with some appreciation of econometric models the elements of econometric modeling are reviewed briefly and in as nontechnical a way as possible. This is followed by a review of national econometric models and regional econometric models. The final section of the text provides numerical examples of how a national econometric model can be used in forecasting and policy analysis. The technical materials on model construction and validation as well as on the use of models in forecasting and analysis is relegated to an appendix.

I. INTRODUCTION

What is Econometrics?

The use of econometric models in economic forecasting and in the design and analysis of public policy has increased substantially over the past several decades. Forecasts based on annual and quarterly econometric models are regularly produced by a number of academic research groups and consulting firms. In addition, these groups also periodically conduct analyses of public policy issues with the use of these models. In view of the increased availability of econometric models for policy analysis, it is desirable for policymakers in responsible government positions to be aware of the types of models that are available and how these models can be used to aid in the formulation of policy recommendations.

This chapter attempts to introduce the basic elements of econometric modeling and to indicate in some detail the types of models that have been developed over the past several decades. The remainder of this section is devoted to a brief, non-technical overview of

Preceding page blank

Econometric Models For Policy Analysis
Discussed in Chapter

Summary Table

Model/Decision Area	Characteristics (Econometric Models)
National Econometric Models	
Bureau of Economic Analysis Model	Ready access to preliminary data
Brookings Model	One of the first attempts at large-scale modeling; sectoral disaggregation
DHL III Model	Large-scale, academic model
DRI Model	Available on a time-shared basis
Fair Model	Primarily a forecasting model relying on anticipations data
Federal Reserve Bank of St. Louis	Small-scale, monetarist model
The MPS Model	Extensive monetary sector
Wharton Mark III	Interaction with industrial subscribers in forecast preparation
Wharton Mark III Anticipations	Extensive use of anticipations data
H-C Annual Model	Designed to study structural change
Wharton Annual Model	I/O table used to derive industry projections
Liu-Hwa Monthly Model	Monthly model limited to some extent by data availability
Regional Econometric Models	
Alaska	Quarterly forecasting and policy analysis model
California (Burton and Dyckman)	Model designed to analyze effects of national policy on the California economy
California (Ratajczak)	Econometric forecasting model
California (Roberts and Wittels)	Forecasting and policy analysis with complete linkages between national economic variables and detailed California economic variables
Southern California (Moody and Puffer)	10-county model using demand components of gross regional product
Hawaii (Chau)	Short-run forecasting model for civilian personal income and employment
Massachusetts (Bell)	Regional forecasting model linked to a national forecasting model
Ohio (L'Esperance, Nestel, and Fromm)	Forecasting and policy analysis model
Puerto Rico (Dutta and Su)	Structural model with emphasis on relationship to U.S. economy
Puerto Rico (Stahl)	Sectoral model to examine productivity trends and relationship of these to domestic investment, trade and income
Other Regional Models	
Lakshmanan and Lo	Regional model for assessment of air pollution abatement programs

econometrics. This introduction includes a discussion of what the term *econometrics* means and how econometrics differs from *industrial dynamics* as well as what these areas have in common. The general principles of econometric model construction are briefly reviewed. The general uses of econometric models are summarized and the section concludes with a concise outline of the types of econometric models that are currently available.

Sections II and III which follow are concerned specifically with the development of econometric models for national and regional economic analysis. The format of each of these sections is the same. Included in each section is a discussion of the structure of national (regional) models, the ways in which national (regional) models have been used, and a short survey of national (regional) models that are currently available.

In section IV, alternative uses of econometric models are considered. Three areas of particular interest to policymakers are singled out for extended discussion:

1. the use of models in forecasting the future;
2. the use of models to investigate the impact of alternative policies; and
3. the use of models to design optimal policies.

Primarily for expository purposes, a recent forecast of the Michigan Quarterly Econometric Model of the U. S. Economy is reviewed. In addition, the results of several policy experiments are discussed. The qualifications to which these numerical results are subject are discussed in some detail so that the policymaker will have an appreciation of what can and cannot be expected from an econometric model.

The Appendix to this chapter includes a technical discussion of the construction and use of econometric models. This material is much more mathematical than the text and should be read only by those familiar with econometric modeling who wish to gain a further appreciation of the technical aspects of model construction, validation, and use in forecasting and policy design.

What is Econometrics? Econometrics is concerned with providing empirical content to economic theory. Economic theory is in turn concerned with the development of concepts and relationships which can be used to describe and interpret observable events. In this sense, econometrics can be thought of as a quantitative approach to the study of economic behavior.

The emphasis in econometrics is on measurement. Two types of measurement are involved: observation of basic data and the estimation of unobserved parameters. The basic data with which the econometrician operates are predominantly non-experimental data which are routinely collected in the course of economic transactions. It is frequently impossible for the economist to generate experimental data which provides the basis for empirical work in the "hard" sciences. This necessitates the development of research strategies in the social sciences and economics in particular which differ from those used in the physical sciences.

An example may be useful to clarify the relationship between theory and measurement in econometrics. To keep the example simple, consider an econometric model of commodity price determination. The policymaker may wish to know the mechanism by which prices are determined in a particular commodity market so that he can introduce policies to stabilize price fluctuations over time. Presented with this problem the econometrician would first turn to economic theory to specify a set of functional relationships that characterize the market. That is, the econometrician would recognize at the outset that a system of relationships is required to describe the market.

A minimal system might involve only three relationships: a demand function, a supply function, and a market clearing equation. More complicated models could easily be constructed but this three-equation system is sufficient to illustrate the basic points and yet simple enough to be manageable. The demand function might be written as

$$D = D(p, q, y, t) \qquad (1)$$

where D denotes the quantity of the commodity demanded by consumers per month, p is the unit price of the good, q is the price of competing goods, y is consumer income, and t is the unit tax on the commodity. The economic theory which underlies this relationship imposes certain conditions on this function. If the commodity is a normal good (as opposed to an inferior good), then a price increase will result in a reduction in quantity demanded, while an increase in income will result in an increase in the quantity demanded. If the price of competing goods increases, the quantity demanded also increases as consumers switch from the higher to the lower-priced commodity.

If the policymaker wished to discourage the consumption of this commodity or to raise revenue by levying a tax on the commodity, it would be necessary to determine empirically the demand function. Only then would it be possible to determine the revenue that could be raised by the imposition of a tax on the commodity. However, it is readily apparent that the demand function alone is not sufficient to provide this information. An increase in the unit tax on the commodity would reduce the quantity demanded by consumers by increasing the effective price. But the process of adjustment is not likely to stop here. The reduction in demand is likely to have some impact on suppliers and their response must also be considered.

This indicates that it is necessary to specify a system of relationships to avoid drawing misleading implications from the analysis. Economic theory again suggests that the supply function is of the form

$$S = S(p, x) \qquad (2)$$

where S is the quantity supplied per month, p is the unit price received by suppliers, and x is an uncontrolled or exogenous variable (i.e. a variable that is taken as given rather than explained by the model) such as the weather that affects production and hence the supply of the commodity. It is generally true that the quantity supplied is an increasing function of the price received by the supplier. In order to close the system, an equilibrium or market clearing equation is required. For this model, it is assumed that prices adjust instantaneously so that the quantity demanded is always equal to the quantity supplied.

$$D = S \qquad (3)$$

These three equations, (1)–(3), determine the equilibrium values of p, D, and S given the values of q, y, and t.

In order to see exactly how this model works, it is convenient to consider linear approximations to equations (1) and (2).

$$D = \alpha_0 - \alpha_1(p + t) + \alpha_2 q + \alpha_3 y \qquad (1')$$

$$S = \beta_0 + \beta_1 p + \beta_2 x \qquad (2')$$

The way in which the demand equation is written indicates that the buyer pays an effective price of p + t per unit but the supplier receives only p, the difference being the tax payment. If (1') and (2') are solved simultaneously with the equilibrium condition (3), the resulting equilibrium price is

$$\bar{p} = (\alpha_1 + \beta_1)^{-1}(\alpha_0 - \beta_0 + \alpha_2 q + \alpha_3 y - \alpha_1 t - \beta_2 x) \qquad (4)$$

In order to determine the impact of a tax change, the derivative of the equilibrium price $\bar{p}$ with respect to the tax rate t is computed. This yields

$$d\bar{p}/dt = -\alpha_1(\alpha_1 + \beta_1)^{-1} \qquad (5)$$

which is seen to involve parameters from both the demand function and the supply function. In addition, the change in tax receipts can be determined. If R denotes tax receipts defined by

$$R = Dt \qquad (6)$$

then

$$\begin{aligned} dR/dt &= D + t dD/dt \\ &= D + t(dD/d\bar{p})(dp/dt) \\ &= D + t\alpha_1^2(\alpha_1 + \beta_1)^{-1} \qquad (7) \end{aligned}$$

which again depends on the parameters of both the demand and supply functions.

The point of this example so far is that the policymaker often deals with problems which can be analyzed only by using a systems approach. It is necessary to consider both the primary and secondary effects of a policy change for the feedback effects may be substantial and may lead to unintended results. This point is, of course, not novel. However, a somewhat more subtle point that is equally important is that, in general, it is not possible to obtain from empirical data adequate estimates of the parameters of the demand function (1) or (1′) without regard to the supply function (2). What this means is that even if the policymaker were interested in only one aspect of a complicated system, empirical results require an examination of the complete system.

This problem, frequently referred to as the identification problem in econometrics, is illustrated in Figure 1. This figure contains plots of the linear demand and supply curves of the commodity market model. The downward sloping demand curve reflects the fact that the quantity demanded decreases with price increases. The upward sloping supply curve reflects the fact that higher prices induce a larger supply of the commodity. The intersection of these two curves determine the equilibrium quantity $\overline{D}$ and the equilibrium price $\overline{p}$. If a set of price and quantity measurements are available, these data may be used to estimate the parameters of the demand curve. Of course observations on the tax rate, the price of competing goods, and income are also required for these variables determine the position of the demand curve. Given these observations, we might be tempted to fit the demand function (1′) by least-squares methods; that is, to determine α_0, α_1, α_2, and α_3 to minimize

$$\sum_{j=1}^{N} e_j^2 = \sum[D_j - \alpha_0 - \alpha_1(p_j + t_j) - \alpha_2 q_j - \alpha_3 y_j]^2 \qquad (8)$$

This is simply the sum of squared errors since

$$e_j = D_j - \alpha_0 + \alpha_1(p_j + t_j) - \alpha_2 q_j - \alpha_3 y_j$$

is this error that arises from the use of $\alpha_0 - \alpha_1(p_j + t_j) + \alpha_2 q_j + \alpha_3 y_j$ to predict the quantity demanded D_j in period j.

A potential problem with this procedure is illustrated in Figure 2. Suppose that the data that are available were observed during a period in which the position of the supply curve was fixed but the demand curve was shifting because of tax-rate changes, income changes, and changes in the prices of competing products. The resulting observations all lie on the supply curve in Figure 2. These data would be appropriate for the estimation of the supply curve but not the demand curve. The opposite case is shown in Figure 3. In this situation the demand curve is stationary and the supply curve is the volatile element. These data would permit the estimation of the demand curve but not the supply curve.

The sets of observations depicted in Figures 2 and 3 are polar extremes. Neither is likely to be observed in reality for both the demand and supply curves are likely to shift from time to time. One of the basic problems in econometrics is to use such data to

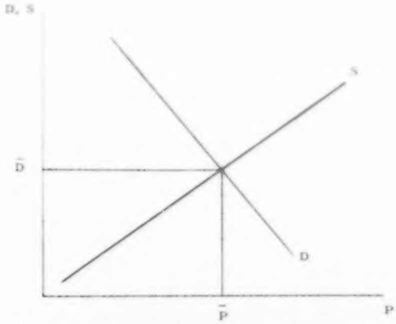


FIGURE 1—*Commodity Market Demand and Supply Curves*

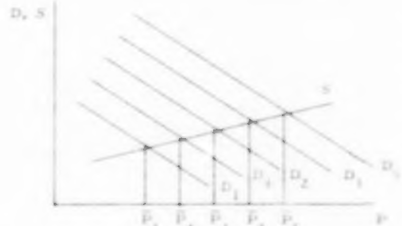


FIGURE 2—*Observations Generated by Shifts in the Demand Curve*

estimate all of the parameters of the interdependent system. This would not be a serious problem if controlled experimentation were possible for then the type of sample data that is required for the estimation problem could be generated. However, as mentioned previously, controlled experimentation is generally not possible in the social sciences. Hence, the econometrician has devised methods to analyze nonexperimental data generated by simultaneous equation systems. Several of these procedures are described below. The point to be emphasized here is that single-equation methods are generally not appropriate since the variables are simultaneously determined. Although this may seem like an obscure technical point, it is necessary to underscore it since many otherwise excellent studies use an inadequate estimation methodology.

In summary, econometrics can be thought of as a branch of statistics concerned with the analysis of non-experimental observations on the variables of an economic system. By way of concluding this brief discussion of econometrics, it may be useful to compare econometrics with industrial dynamics. Viewed as research methods, the differences among these areas would appear to be primarily differences in emphasis rather than substance. Industrial dynamics as described by Forrester [7, 8], for example, is concerned with the construction of complex dynamic models that describe physical or social processes. The complexity of the models that have been developed in this area lead to two serious problems. First, typically the data required to estimate the parameters are not available. Even if they were available, the functional forms that are postulated are such that parameter estimation is difficult if not impossible. The second problem associated with the complexity of industrial dynamics models is that analytical solutions are difficult to obtain. It is therefore often necessary to resort to numerical procedures to analyze the resulting models. This is not a major difference between econometric and industrial dynamics models however. The critical difference is the relative importance of parameter estimation in the approach to modeling.

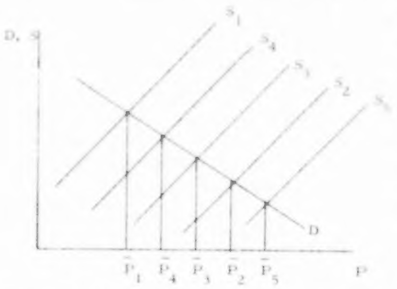


FIGURE 3—*Observations Generated by Shifts in the Supply Curve*

The estimation problems inherent in industrial dynamics models are often quite severe. However, this is not a serious problem if one of the basic postulates of the industrial dynamics methodology is satisfied. Specifically, it is asserted that the behavior of complex systems is not critically dependent on the parameter values. If this is true, that is, the same conclusions are obtained for wide ranges of parameter values, then the problem of parameter estimation is not so serious. The emphasis in econometric modeling is on parameter estimation. In essence, the postulate that the system behavior is insensitive to parameter variation is rejected, at least initially. This frequently leads the econometrician to consider models which offer some promise for parameter estimation and to reject models that are too complicated to allow estimation to proceed with the data that are available.

Uses of Econometric Models. The distinguishing features of econometric modeling are the use of economic theory in the specification of the economic system and the use of statistical theory to estimate the parameters of the specified model using non-experimental observations. The process of model construction, including estimation and validation, will be discussed at length in the Appendix. For purposes of exposition, it is assumed that this stage of the analysis has been completed. This section discusses briefly the ways in which econometric models can be used by decision makers.

Returning to the commodity market example, the parameters of the linear model (1′)–(3′) are now assumed to be known. If a_i denotes the estimated value of α_i and b_i denotes the estimated value of β_i, the empirical counterparts of (1′) and (2′) are

$$D_j = a_0 - a_1(p_j + t_j) + a_2q_j + a_3y_j \qquad (1e)$$

$$S_j = b_0 + b_1p_j + b_2x_j \qquad (2e)$$

where the subscripts refer to time periods. It might be necessary to introduce dynamic adjustments to improve the accuracy of the model if, for example, suppliers do not respond instantaneously to price changes. This refinement will not be introduced for the sake of simplicity.

This model can now be used in two ways: prediction of future values of the system and analysis of policy changes. The empirical system (1e), (2e), and (3) can be used to predict in advance the commodity price and the quantity sold. The solution for the equilibrium price is

$$\hat{p} = (a_1 + b_1)^{-1}(a_0 - b_0 + a_2q_j + a_3y_j - a_1t_j - b_2x_j) \qquad (4e)$$

Predictions are obtained by substituting predicted values for q_j, y_j, t_j, and x_j in this expression and solving for $\hat{p}_j$. The resulting prediction is a conditional prediction in the sense that is is the price that is expected to prevail if the predicted values of the independent variables are realized. The conditional prediction given by (4e) is subject to error from three sources. First, if the independent variables deviate from their projected values, this will introduce an error. Second, even if the independent variables are forecast correctly, any errors in the coefficient estimates will lead to prediction error. Third, the fact that the model is not exact will introduce a further error in the prediction.

This last source of error is not apparent from the overly simplified way in which the model has been presented. In point of fact, the system (1)–(3) is considered to be not an exact system but rather a stochastic system. Hence a more realistic representation would be

$$D = D(p, q, y, t, u) \qquad (1s)$$

$$S = S(p, x, v) \qquad (2s)$$

$$D = S \qquad (3s)$$

where u and v are random disturbances. These random disturbances arise because the quantity demanded is not an exact function of p, q, y, and t. Similarly, the quantity supplied does not depend exclusively on p and y but also on other influences which are lumped together in the random variable denoted by v. In going from the nonlinear formulation (1s)–(2s) to the linear formulation (1′)–(2′) another source of error—misspecification of the form of the relationships—may be involved. In any event, the linear model should have been written as

$$D = \alpha_0 - \alpha_1(p + t) + \alpha_2q + \alpha y + u \qquad (1's)$$

$$S = \beta_0 + \beta_1p + \beta_2x + v \qquad (2's)$$

It is now apparent that the solution of (1′s), (2′s), and (3) for p_s will depend on u and v, i.e.,

$$\bar{p}_s = \bar{p} + (\alpha_1 + \beta_1)^{-1}(u - v) \qquad (4's)$$

where $\bar{p}$ is given by (4). Hence the error in using $\bar{p}$ to predict the actual price $\bar{p}_s$ is $(\alpha_1 + \beta_1)^{-1}(u - v)$. The other two sources of error result from the use of $\hat{p}$ to estimate $\bar{p}$.

The analysis of these sources of error is an important part of any applied econometric study. A lack of data often prevents a detailed analysis of prediction error but as discussed in detail below, prediction error analysis is an extremely important part of the process of validation of a model. Recent work of Howrey [14] and Dhrymes, et. al. [5] suggests that three types of tests are of special importance in connection with validation of a model. These are

- Tests of data conformity,
- Comparisons of alternative models, and
- User-oriented criteria of success.

The statistical aspects of these tests are described in detail in the Appendix. It is perhaps sufficient to note at this point that these tests all depend on forecast errors in one way or another.

The second general way in which the model can be used is in the study of policy issues. For example, with the empirical model, the policy-maker is able to estimate the magnitude of the response of price to a change in the tax rate. That is, the derivative calculated in (5) can now be assigned a value equal to $-a_2(a_1 + b_1)^{-1}$. This is only an estimate since the true values of the parameters are not known. How good an estimate of the policy impact this yields depends on the extent to which the model is a valid representation of the system and how close the parameter estimates are to their true values.

In more complicated models, alternative policies can be compared in this way. Suppose, for example, that the independent variable x in the supply equation represents a unit subsidy to the supplier. The supplier now receives p + x for each unit produced and sold on the market and the consumer pays p + t for each unit purchased. The model can be used to compare a subsidy program with a tax-reduction program. For example, suppose that it is desired to increase the production of a commodity that is currently taxed at the rate t. The alternative policies that are under consideration include a tax reduction or a subsidy to producers. A tax reduction in the amount Δt will result in an estimated price reduction of

$$\Delta\bar{p} = a_2(a_1 + b_1)^{-1}\Delta t$$

An increase in the subsidy by Δx will result in an expected price reduction in the amount of

$$\Delta\bar{p} = b_2(a_1 + b_1)^{-1}\Delta x.$$

Which of these two policies results in the larger price reduction will depend on the values of the parameters α_2 and β_2 of the model. The estimates a_2 and b_2 can be used to provide evidence on this issue.

These examples illustrate the use of the model to compare different *ad hoc* policies. If an objective function is introduced, then the model, together with the objective function, can be used to determine the best of several alternative policies. Returning to the taxation/subsidy comparisons, the policymaker might introduce as the objective function the change in output that results from the policy and the cost of the policy. If two policies achieve the same expansion of output but one policy costs more than another, the least-cost policy is preferable. The objective function thus permits the ranking of alternative policies on the basis of the results that they produce and the model provides the basis for the estimation of the results of different policies.

This evaluation procedure can be carried one step further. Suppose that the policymaker wishes to increase output by a specified amount but to do so in the least cost way. The policymaker has two variables that can be controlled, the tax rate and the subsidy rate. The problem is to find the optimal combination of these two rates that achieves the output goal and minimizes cost. This optimization problem illustrates the use of econometric models in the selection of optimal policies. This subject will be described in greater detail in the Appendix.

Types of Econometric Models. The field of econometrics consists of two types of activities: econometric theory and applied econometrics. Econometric theory is concerned with the development of statistical techniques which are used in estimation and hypothesis testing. Applied econometrics is the application of these techniques in concrete modeling situations. It is difficult to construct a complete catalogue of econometric models because econometric methods have been applied to so many diverse problems. Many of the models that have been discussed in other chapters of this book could quite properly be called econometric models.

Broadly defined, an econometric model is any model dealing with economic relationships in which statistical methods have been employed to estimate unknown parameters or to test hypotheses. This definition is clearly so broad that it encompasses all but the most theoretical of economic models. Perhaps a more useful way to categorize econometric models is in terms of the subject matter with which they are concerned. In this regard, a first broad distinction could be made between microeconomic models and macroeconomic models. Microeconomic models are concerned with the behavior of individuals or small groups of individuals. Thus models dealing with the expenditure decisions of a household, the investment decisions of a firm, or the price of a particular commodity would be microeconomic models. Macroeconomic models are concerned with behavior in the aggregate. The main variables of these models are aggregate income, the unemployment rate, wage and interest rates, and the price level. Models which attempt to describe the national economy or the economy of a state or region are examples of macroeconomic models.

The models that are carefully reviewed subsequently are national and regional macroeconomic models. There are several reasons for choosing this area to review. First, by far the largest amount of coherent, systematic model-building effort has been devoted to macroeconomic systems. Hence there is a fair amount of agreement on the general structure of such models and a substantial number of applied results for national econometric models. Second, methods of regional economic analysis are advancing quite rapidly and several operational regional models have recently been constructed. This is an area which is likely to experience significant breakthroughs in the near future in response to urgent needs of economic policy-makers. Finally, national and regional models lend themselves quite naturally to a decision-making context. Moreover, the development of these national and regional models as decision-making tools would benefit enormously from a more direct input from policy-makers.

Confining specific attention to national and regional models means that several important areas of applied work are not discussed. In particular, commodity or industry models are not reviewed in any detail here. The expository example of the introduction should give the reader a good feel for what is involved in the construction of a commodity market model and how it can be used for decision-making purposes. For a more detailed discussion of this type of model, the interested reader is referred to the recent book by Labys [91].

Within the broad area of national and regional economic models, it is possible to make the following distinctions.

- Long-term/Short-term Models
- Forecasting/Policy Models
- Large-scale/Small-scale Models

Short-term models are concerned primarily with near-term prediction and policy analysis. There are several national models that produce eight-quarter forecasts each quarter. The methods of construction and implementation of these short-term models are somewhat different from those used to build models which are intended to provide ten or twenty-year projections.

It is also possible to distinguish between models primarily designed for forecasting and those designed for policy analysis. Forecasting models, particularly short-term forecasting models, can make use of anticipations data, that is, data obtained from preliminary surveys of household and business intentions. Less attention is devoted to the introduction of detailed policy variables and the channels through which the policy variables operate. Policy models, on the other hand, must necessarily be concerned with this aspect of the system.

The scale of the model can often be varied almost continuously depending upon how much detail is required and the amount of resources available for the construction and operation of the model. National models range in scale from five or six equations to two-hundred or more equations. Slightly different strategies are involved in the construction of very large-scale models than in smaller models. The problem of coordination as it affects the logical consistency and completeness of the model becomes acute if a number of investigators are involved in the development of the model. However, the overall approach to the model is not critically dependent on the scale of the model.

II. NATIONAL ECONOMETRIC MODELS

In order to gain an appreciation of the potential uses of national economic models, it is necessary to understand the structure of these models. A fairly elementary expository model is introduced to illustrate the basic relationships of a macroeconometric model. This expository model is then used to indicate the uses of such a model for policy analysis. This section is concluded with a brief survey of operational econometric models of interest to policy-makers.

An Expository National Econometric Model. The structure of the most complicated econometric model of the national economy can be grasped fairly easily once this expository model is understood. The way in which large-scale, operational models differ from this system will be explained after the model has been introduced. It should be noted at the outset, however, that most econometric models are dynamic in the sense that they include lagged values of the dependent and independent variables. In the interests of economy of presentation, this expository model will, at least initially, suppress the dynamic elements in order to concentrate on the general form of the relationships.

The reader will recall that the simple commodity market model introduced previously consisted of three equations: a demand equation, a supply equation, and a market clearing or price determination equation. A national macroeconomic model typically has the same three blocks of equations: a set of demand equations, a set of supply equations, and a set of price equations. The demand equations of the model attempt to explain broad expenditure aggregates such as personal consumption expenditures, gross private domestic investment expenditure, and net foreign investment. These aggregates closely follow the national income and product account data which are typically used to estimate the parameters of the models.

The following equations provide an example of the block of demand equations of a national econometric model.

$$C = C(y, r, m) \qquad (9)$$

$$I = (X, r, K) \qquad (10)$$

$$X = C + I + G \qquad (11)$$

The symbols used in these equations are defined as follows.

C : real (constant-dollar) personal consumption expenditure
y : real disposable income
r : the interest rate
m : real balances held by households

I : gross private domestic investment
K : the stock of capital
X : real aggregate demand
G : real government expenditure

Personal consumption expenditure is assumed to be a function of real disposable income, the interest rate, and the stock of real balances held by households. This equation is based on the idea that consumption expenditure depends on the resources of households as reflected in their income y and wealth m. In addition, the opportunity cost of consuming and hence not saving, as reflected in the interest rate, may also have an impact on the level of consumption expenditure.

The investment function states that the level of investment is a function of the level of aggregate demand, the cost of capital as measured by the interest rate, and the current stock of capital. If the current stock of capital is large relative to aggregate demand so that there is excess capacity, then investment spending will be low. On the other hand, if the current stock of capital is low relative to current aggregate demand, there will be pressure to expand production facilities and investment spending will be influenced to some extent by the availability and cost of funds. Hence the interest rate is included in this equation to capture this effect.

The final equation of this block simply sums the individual demand components to arrive at the aggregate demand for output. The level of government expenditure, as in most models, is assumed to be exogenous. The model does not attempt to explain variations in government spending; rather, these are taken as given.

The supply block of this model consists of four equations.

$$X = X(K, L) \qquad (12)$$

$$L^f = L(w/p) \qquad (13)$$

$$K = K_{-1} + I - D \qquad (14)$$

$$D = \delta K \qquad (15)$$

The new symbols which appear in this block are as follows.

L : the level of employment

L^f : the full-employment supply of labor

w : the wage rate

p : the price level

D : the level of capital consumption

The first equation of this block is the production function which indicates how capital and labor are combined to produce the output X. The second equation states that the supply of labor depends on the real wage rate. The last two equations determine the supply of capital. The capital stock available for current production is equal to the stock inherited from last period plus the amount of gross investment less retirements. Capital consumption is assumed to be proportional to the stock of capital.

The block of price equations for this model can be written in the following way.

$$p = p(wL/X) \qquad (16)$$

$$\Delta w = w(L^f - L) \qquad (17)$$

$$m = m(y, r) \qquad (18)$$

The price equation (16) is based on the notion that prices depend on unit labor costs wL/X. The wage adjustment equation indicates that wages adjust in response to unemployment since $L^f - L$ is the difference between the full-employment labor force and the amount of employment currently available. The final equation is a very simple characterization of the monetary sector in which the interest rate adjusts to equilibrate the demand for money m(y, r) with the supply of money m which is in turn determined by the banking system and the financial sector of the economy.

Several definitions and accounting identities are needed to close the system.

$$Y = pX - Tx + Tr - pD \quad (19)$$

$$y = Y/p \quad (20)$$

$$m = M/p \quad (21)$$

The new symbols introduced in these identities are defined as follows.

Y : disposable personal income
M : the stock of money
Tx : federal, state, and local tax payments
Tr : federal, state, and local transfer payments to households

The first of these identities indicates that disposable income is equal to the market value of all output produced less taxes paid plus transfers to households less capital consumption allowances. The other two identities indicate how the real and nominal values are related.

This system of 13 equations describes how the 13 endogenous variables C, I, X, L, L^f, K, D, p, w, r, Y, y, and m are determined. The exogenous variables of the model are G, Tx, Tr, and M. Thus given the values of these exogenous variables, it would be possible in principle to solve for the 13 endogenous variables of the model.

The implementation of this model requires that the general functional relationships be specified in a form in which the unknown parameters can be estimated. This involves two things. First, the general functional form must be replaced by a specific, parametric relationship. Using the consumption function as an example, this function might be written as

$$C = c_0 + c_1 y + c_2 r + c_3 m + c_4 C_{-1} + u \quad (22)$$

In this form, the unknown parameters $c_0, c_1, \ldots, c_4$ could be estimated from observations on C, y, r, and m. The disturbance term u is added to the equation to indicate that the equation is not exact but rather a stochastic relationship. This error term arises because the functional form may not be correct, variables that are relevant to the determination of personal consumption expenditure may have been omitted from the equation, or consumption behavior of households may be inherently random rather than deterministic. The specification of the properties of this error term is the second aspect of the specification phase that is required before estimation can proceed.

Each of the equations in the model which involves unknown parameters must be specified in this way. Of course, the identities need not be modified since they do not involve unknown parameters and hold exactly. In the above system there are eight stochastic equations: (9), (10), (12), (13), (15), (16), (17), and (18). The remaining five equations, (11), (14), (19), (20), and (21) are identities.

Uses of a National Econometric Model. Once the parameters of the model have been estimated, the model can be used for forecasting and policy analysis. In connection with the system described above, forecasts of all of the endogenous variables of the model, i.e., variables that are explained by the model, can be obtained once the exogenous or independent variables are predicted. Thus in order to forecast the level of gross national product next period, it is necessary to forecast the level of government spending, tax collections, transfer payments, and the money stock. Once these are given, the system can be solved to obtain a forecast of the level of output next period. It should be emphasized that this forecast is a conditional forecast in the sense that it depends on projected values of the exogenous variables. Forecast errors arise from three sources: the use of estimates in place of the unknown coefficients of the system, the disturbance errors in the stochastic equations, and errors in forecasting the exogenous variables.

In addition to the use of a national econometric model as a forecasting device, it can be used to analyze the impacts of alternative policies. For example, modifications of the tax system which affect Tx in the system can be investigated. Changes in welfare programs which alter the value of Tr are likely to have an impact on income, employment, and prices and these effects can be estimated using the model. Finally, changes in government spending which affect G or changes in monetary policy which result in changes in M can be investigated. After an investigation of a number of alternative policy changes, the policy which seems most appropriate can be chosen by the decision-maker.

Operational National Econometric Models. The simple expository model introduced above is intended to provide an overview of the structure of an operational econometric model. The operational models that are currently in use are typically much larger than this simple model. The block of demand equations typically contains many more equations than those shown in the expository model. For example, personal consumption expenditure is disaggregated by type of expenditure to give consumption of durables, consumption of non-durables and consumption of services. Durables consumption is sometimes further subdivided into automobile expenditures, expenditures on household goods, and expenditures on other durables. Since the prices of these different components do not always move together price equations must be introduced for each component of consumption expenditure.

The single investment function (10) is typically replaced by a set of investment functions corresponding to different types of investment expenditure categories. For example, a separate equation might be included for investment in plant and equipment, residential construction, and investment in inventory. These investment equations could be further disaggregated by industrial type. Again since prices of these different types of investment goods do not move together, price equations would be introduced for the different components.

In most models export and import equations are added to the system. These additions result in the modification of the identity (11) to

$$X = \sum C_i + \sum I_i + Ex - Im + G \qquad (11')$$

where C_i denotes the demand for consumption of type i, I_i is investment demand for the i^{th} type of investment good, Ex denotes exports, and Im denotes imports.

The supply block can be expanded by introducing production functions for different types of broad product groups. Similarly, the production functions can be modified to include different types of labor. The labor supply and capital supply functions can be expanded in the same way.

It has already been noted that the introduction of consumption and investment components necessitates an expansion of the block of price equations. If different types of labor are introduced, then a corresponding set of wage adjustment equations can also be introduced in place of the single wage adjustment equation (17). The rudimentary monetary sector embodied in (18) is typically expanded to explain long-term and short-term interest rates, the levels of demand and time deposits and the response of financial institutions to changes in the discount rate and open market operations conducted by the federal reserve system.

These modifications result in rather large and complex nonlinear systems of equations. Although they are used in exactly the same way as the simpler models that have been used here for expository purposes, solution by numerical methods is frequently required. The size and complexity of these models means that the operation of these models is a very specialized activity.

An excellent tabular survey of a number of econometric models is given by Nerlove [56]. This survey is somewhat dated now, but it provides an overview of the state of the art as of about 1965. A recent paper by Fromm and Klein [46] provides some comparisons of the properties of the models that are most readily available and widely used today. The models included in this survey include the following.

- The Bureau of Economic Analysis Model, A. Hirsch, M. Leibenberg, and G. Narasimham [48, 53]
- The Brookings Model, The Brookings Institute, G. Fromm, L. Klein, and G. Schink [39, 40, 47]
- The DHL III Model, University of Michigan, S. Hymans and H. Shapiro [49]
- The Data Resources Inc. Model, O. Eckstein and E. Green
- The Fair Model, Princeton University, R. Fair [44]
- The Federal Reserve Bank of St. Louis Model, L. Anderson and K. Carlson [33]
- The MPS Model, University of Pennsylvania, A. Ando, F. Modigliani, and R. Rasche [34]
- The Wharton Mark III Model, University of Pennsylvania, F. G. Adams, V. J. Duggal, G. Green, L. Klein, and M. McCarthy [55]
- The Wharton Mark III Model, Anticipations Version [32]
- The H-C Annual Model, Stanford University, B. Hickman and R. Coen
- The Wharton Annual Model, University of Pennsylvania, R. Preston
- The Liu-Hwa Monthly Model, Cornell University, T. C. Liu and E. C. Hwa

With the exception of the last three models, all are quarterly models. It is difficult to provide a concise, up-to-date description of these models because they are all in the developmental stage in the sense that minor modifications are made periodically. A complete description of the current version of the model can often be obtained by writing directly to the model builder.

The paper by Fromm and Klein referred to above gives comparative figures on a number of characteristics of these models. These characteristics include

- Sample-period solution errors
- Post-sample-period solution errors
- Dynamic tax multipliers
- Dynamic government expenditure multipliers

These results provide a basis for making some judgments about the accuracy that econometric forecasts can be expected to achieve. In addition, the differences in the dynamic multipliers obtained from the different models indicates the range of the uncertainty which accompanies such estimates.

Another set of model comparisons is contained in the paper by Cooper [37]. These results have generated a certain amount of controversy and have been responsible, at least in part, for a substantial amount of research activity which is only now being completed. Of special note along these lines is the recent paper by Nelson [96] which compares the forecast accuracy of the MPS model with naive autoregressive models. This methodology of model evaluation has been examined in some detail by Howrey, Klein, and McCarthy [16]. A discussion of the predictive accuracy of Wharton forecasts is contained in Green and Klein [11].

III. REGIONAL ECONOMETRIC MODELS

The demand by policymakers for more detail at the state and local level has led to the development of regional econometric models. Most of these models have been constructed along the same lines as the national macroeconometric models discussed in the preceding section. These regional models are usually concerned with state aggregates. This is perhaps not the ideal level of aggregation but what little data exists are usually collected on a statewide basis.

This review of regional models follows the same format as that of the preceding section. The structure of a typical regional model is introduced first. Ways in which regional models can be used for policy analysis are then described. The review concludes with a few comments on operational regional models.

An Expository Regional Econometric Model. In an excellent survey paper, Klein [76] has outlined the structure of a representative regional econometric model. The actual models that have been constructed depart from this structure in some areas to account for the peculiarities of the area under investigation. However, the general structure proposed by Klein is followed to a large extent and is extremely useful in understanding the nature of regional econometric models.

The usual approach that is pursued in regional econometric modeling is to follow the outline of a national model fairly closely. Some adjustments are, of course, required by the nature of economic activity in the region. What this means in terms of regional modeling is that aggregate production and employment is explained by the model through an aggregate demand equation. The alternative approach is to develop a set of interrelated models for each of the major industries of the area and then to aggregate these industry relationships. This approach has not been followed primarily because of lack of detailed data that this approach would require. However, another reason for not pursuing this approach to regional modeling is the success that has been attained by national models that are constructed from aggregate data.

The block of demand equations of a representative regional model can be written as follows.

$$C = C(Y/pc^*) \qquad (23)$$

$$I = I(X, r^*, k) \qquad (24)$$

$$G = G(Tx, N, r^*) \qquad (25)$$

$$Ex = Ex(X^*, pe/pm^*) \qquad (26)$$

$$Im = Im(X, p/pm) \qquad (27)$$

$$pX = pc\,C + q^*I + G + G^* + pe\,Ex - pm\,Im \qquad (28)$$

The variables included in this block are defined in the following way. An asterisk denotes that the variable is a national aggregate. Otherwise the variable is a regional variable.

C : regional personal consumption expenditure
Y : disposable personal income
pc^* : national index of consumer prices
I : gross private domestic investment
X : gross regional product
r^* : the interest rate
K : the capital stock
G : state and local government expenditure on goods and services
Tx : state and local taxes
N : regional population
Ex : exports of goods and services
X^* : gross national product
pe : price of exports
pm^* : the price of imports
Im : imports of goods and services
p : gross regional product deflator
q^* : cost of investment goods
G^* : federal government expenditure on regional goods and services

The reader will recognize that these functions are similar to those given previously for the national model. The important difference is the mixture of national and regional variables in this model and this requires further comment.

The regional consumption function states that consumption expenditure in the region is primarily a function of disposable personal income of households in the region deflated by the national price index of consumer goods. The reason for the use of the national index is that consumer purchases in any region will usually contain a large number of items produced outside the region. It is therefore more appropriate to use a national price deflator than a regional price deflator. The regional investment function is of precisely the same form as the national investment function. The use of a national interest rate r^* is appropriate here since business firms can usually obtain funds in a national market. They are not confined to borrowing in local markets.

A major difference between national and regional models is that in regional models it is thought to be appropriate to consider state and local government expenditure as an endogenous variable to be determined within the system. State and local expenditure is usually closely tied to tax receipts, population, and the cost of borrowing funds. This link between receipts and expenditures is much stronger at the state and local level than it is at the national level.

The fact that regional economies are much more open than national economies means that the export and import functions play an especially important role in regional models. The export function postulates a relationship between regional exports and gross national product, the regional export price, and the price of imports from all other regions. The idea is that a region can expect to export more to other regions during periods of prosperity. There are two reasons for this. High levels of activity will generate demands for inputs. High levels of activity also generate high incomes and hence lead to bouyant demand. In addition to expanded activity, exports are higher if export prices are low relative to domestic prices in other regions. Hence the introduction of the price ratio pe/pm^* in the export equation. The import equation is based on the same reasoning. Imports are expected to vary positively with regional production and income and directly with the price ratio p/pm^*. An increase in p/pm^* means that domestic goods are more expensive relative to imported goods and this would be expected to stimulate import demand.

The final regional demand equation aggregates the individual regional demand components. The only completely exogenous variable in this equation is the federal expenditure on regional goods and services. None of the other components of demand is completely endogenous in the sense that each one depends in part on a national economic variable.

The supply side of the regional model is characterized by four equations.

$$X = X(K, L) \qquad (29)$$

$$K = K_{-1} + I - D \qquad (30)$$

$$D = D(K) \qquad (31)$$

$$L^s = L(N) \qquad (32)$$

The new variables introduced in this block are as follows.

L : regional labor force
D : regional capital consumption
L^s : regional supply of labor

Since these equations are quite similar to those of a national model, detailed comment is not required. It might be noted that a very simple labor supply function is introduced here in which the supply of labor depends on regional population and not on the regional wage rate as in the national model.

The price equations of the regional model are of a simple form.

$$p = p(p^*, w, pm^*) \qquad (33)$$

$$pe = pe(p^*, w, pm^*) \qquad (34)$$

$$w = w(u, u^*, pe^*) \qquad (35)$$

The new variables introduced in this block are the following.

w : regional wage rate
u : regional unemployment rate
u^* : national unemployment rate

These price equations are somewhat *ad hoc* formulations. However, they are similar to the price equations that are employed with considerable success in national models.

A rather detailed tax and transfer system can often be constructed for a regional model. The following system of equations provides an example of such a system.

$$Tx = Txi + Txd \qquad (36)$$

$$Txi = Txi(pX) \qquad (37)$$

$$\mathrm{Txd} = \mathrm{Txd}(\mathrm{w\,L}, \pi) \quad (38)$$

$$\mathrm{Txf} = \mathrm{Txf}(\mathrm{w\,L}, \pi) \quad (39)$$

$$\mathrm{Tr} = \mathrm{Tr}(\mathrm{u}, \mathrm{N}) \quad (40)$$

These symbols have the following meaning.

Tx : total state and local taxes
Txi : indirect state and local taxes
Txd : direct state and local taxes
Txf : federal tax collections in the regional
Tr : regional transfer payments
π : nonwage income

Finally, two identities are needed to close the system.

$$\pi = \mathrm{pX} - \mathrm{qD} - \mathrm{wL} - \mathrm{Txi} \quad (41)$$

$$\mathrm{u} = \mathrm{L}^* - \mathrm{L} \quad (42)$$

This completes a system of twenty equations in the twenty endogenous variables C, I, G, Ex, Im, X, L, K, D, L^*, p, pe, w, Tx, Txi, Txd, Txf, Tr, π, and u. In order to solve for these twenty endogenous variables, the values of the national variables pc^*, r^*, X^*, pm^*, q^*, and u^* are needed as well as the exogenous variables N and G^*.

Uses of Regional Econometric Models. Regional econometric models are used in much the same way as national econometric models are used. In the context of economic forecasting, predictions of the endogenous regional variables can be obtained by solving the regional model with projected values inserted for the exogenous national and regional variables. The use of regional models in conjunction with national econometric models is of relevance in this connection. The national model can be used to generate forecasts of the national variables that are used in turn as inputs to the regional model.

Regional models are particularly useful to the policy-maker in connection with projecting future tax receipts. Unlike the Federal government, State and local governments often are faced with difficulties in borrowing to cover unforeseen deficits. It is therefore important to be able to anticipate potential deficits in State and local government budgets and to make contingency plans to cover these deficits. Regional models can also be used to obtain forecasts of unemployment and the demand for transfer payments through welfare programs.

Regional models can be used in policy analysis to explore the impact of alternative tax and expenditure policies. For example, the impact of revenue-sharing proposals can be investigated using a regional econometric model. One of the variables in the model described above is the level of Federal spending on regional goods and services. The regional impact of changes in Federal spending can be derived with the use of a regional model.

A recent interesting use of regional economic models is provided by Lakshmanan and Lo [95]. They develop a model to predict the cost of air pollution control in a region under alternative control schemes. Since the model is one of the few ventures in this area, it is worth reviewing briefly the basic elements that are involved. The model incorporates four sets of interrelationships, namely,

1. A Keynesian-type regional macro model of the type discussed above;
2. A set of production relationships by industry;
3. A set of regional fuel and electricity demand equations by industry; and
4. A set of relationships linking regional industrial activity with national industrial activity.

The impacts of alternative air pollution abatement policies are determined by assuming that the cost of the products of high-emission industries will increase. This will result in a reduction in the demand for these products and will have regional effects on income and employment that can be traced through the model. Although the model is still in the experimental stage, interesting results on the distributional effects of pollution abatement policies have been derived. For a detailed discussion of the model and the policy results that have been obtained, the reader is referred to Lakshmanan and Lo [95].

Regional modeling is a relatively new venture. The results that have been obtained to date are sufficient to establish regional econometric modeling as a useful research area. However, progress has been relatively slow and the models that have been constructed are analogous to national models. As this area develops, it is likely that new applications will be found for these models.

Operational Regional Econometric Models. The major U. S. regional econometric models have recently been reviewed by Glickman [70] and Chen [64]. The models reviewed by Chen include the following:

- The Alaska Model, G. H. Tuck [90]
- The California Model, R. P. Burton and J. W. Dyckman [62]
- The California Model, D. Ratajczak [84]
- The California Model, B. F. Roberts and G. Wittels [87]
- The Ten Southern California Counties Model, H. L. Moody and F. W. Puffer [79]
- The Hawaii Model, L. C. Chau [63]

Table I

Stated Objectives of the Ten Regional Econometric Models

Model	Stated Objectives of the Model Builder
ALASKA Tuck	Construct an aggregate income model of Alaska for quarterly economic forecasts, 4 to 6 quarters in advance, for policy analysis, and systematic organization of data.
CALIFORNIA Burton Dyckman	The basic purposes of the Phase I version of the model were for economic forecasting and analysis of the effect of national policy upon the California economy. The Phase II version attempted to refine the Phase I model by improving estimation reliability of certain sectors, reclassifying certain industry groupings, and securing better predicting equations through improvement or augmentation of basic data.
CALIFORNIA Ratajczak	Develop an econometric forecasting model of California.
CALIFORNIA Roberts Wittels	The CEEP California model was intended for use by decision makers for systematically projecting images of the California economy and for posing meaningful economic questions. The structure of the model draws complete linkages between national economic policies and detailed California economic variables.
SOUTHERN CALIFORNIA Moody Puffer	Construct an economic model of the 10 southern-most counties of California utilizing the demand components of Gross Regional Product.
HAWAII Chau	Construct an economic model for making short-run forecasts of civilian personal income and employment in Hawaii.
MASSACHUSETTS Bell	Develop a regional econometric model which can be linked to national forecasts and apply this model to Massachusetts.
OHIO L'Esperance Nestel Fromm	Develop a set of social accounts for estimating the gross state product of Ohio. Construct an econometric model describing Ohio's economic behavior which is useful for forecasting and policy analysis.
PUERTO RICO Dutta Su	Construct an econometric model which incorporates all important features of her economic structure, especially her close economic relationship with the U.S.
PUERTO RICO Stahl	Use a model similar to L. R. Klein's model of the Japanese economy to examine trends in labor productivity in various sectors of the economy and the relationship between these productivities, investment, external trade, and the level of income in Puerto Rico.

Source: Dean Chen [64].

- The Massachusetts Model, F. W. Bell [58]
- The Ohio Model, W. L. L'Esperance, G. Nestel, and D. Fromm [78]
- The Puerto Rico Model, M. Dutta and V. Su [66]
- The Puerto Rico Model, J. E. Stahl [88]

Glickman's survey includes the Burton and Dyckman California Model, The Massachusetts Model, and the Ohio Model. In addition, the following models are reviewed by Glickman.

- The Michigan Model, D. B. Suits [86]
- The Northeast Corridor Model, R. T. Crow [65]
- The Philadelphia Model, N. J. Glickman [68]

The potential uses of these models can be inferred, at least partially, from the stated objectives of the model-builders. These objectives have been summarized by Chen for the models in his review and the tabular summary is reproduced in Table I. It is apparent that the objectives of the model-builders are similar and stated only in general terms. The development of a forecasting model for the region is most frequently given as the stated objective of the investigation. Policy analysis is also mentioned specifically by some of the model-builders but no specific policy issues are mentioned.

Table II, also constructed by Chen,

Table II

Type of Data of the Ten Regional Econometric Models

Model	Type of Data	Period Covered
ALASKA Tuck	Quarterly	1960–67
CALIFORNIA Burton Dyckman	Quarterly	1950–62 for all equations which do not contain predetermined defense variables 1954–62 for all remaining equations which do contain exogenous defense variables
CALIFORNIA Ratajczak	Quarterly	1953–1971
CALIFORNIA Roberts Wittels	Quarterly	1961–I to 1971–III
SOUTHERN CALIFORNIA Moody Puffer	Annual	1953–61
HAWAII Chau	Annual	1951–68
MASSACHUSETTS Bell	Annual	1947–62
OHIO L'Esperance Nestel Fromm	Annual	The model was estimated with varied lengths of period for different equations. The maximum length of period covers 1945–64.
PUERTO RICO Dutta Su	Annual	1948–64
PUERTO RICO Stahl	Annual	1947–61

Source: Dean Chen [64].

indicates some of the characteristics of these models. The type of data and the sample period used to estimate the parameters are given for each model. Most of the models have been estimated by ordinary least squares. However, the models for two states, Massachusetts and Ohio, have also been estimated using two-stage least squares.

The scale of these regional models is indicated in Table III which was also compiled by Chen. For each model, the number of stochastic equations and the number of identities are given. The sum of these two numbers is the number of endogenous variables of the model. These models range in size from Stahl's thirteen-equation model of Puerto Rico to the 294-equation model of California devised by Burton and Dyckman.

The final table that has been extracted from Chen's review is Table IV which gives the number of exogenous variables in each model. These are separated into the number of regional exogenous variables and national exogenous variables. The larger the number of exogenous variables, the more difficult it is to use the model for prediction and policy analysis. The reason for this is that to obtain a prediction of the endogenous variables it is necessary to forecast the exogenous variables. The models range from a low of eight exogenous variables

Table III

Number of Endogenous Variables of the Ten Regional Econometric Models

Model	Stochastic Equations	Identities	Total Endogenous Variables
ALASKA Tuck	14	18	32
CALIFORNIA Burton Dyckman	240	54	294
CALIFORNIA Ratajczak	32	32	64
CALIFORNIA Roberts Wittels	33	43	76
SOUTHERN CALIFORNIA Moody Puffer	13	5	18
HAWAII Chau	25	6	31
MASSACHUSETTS Bell	8	6	14
OHIO L'Esperance Nestel Fromm	16	11	27
PUERTO RICO Dutta Su	23	12	35
PUERTO RICO Stahl	11	2	13

Source: Dean Chen [64].

Table IV

Number of Exogenous Variables of the Ten Regional Econometric Models

Model	Regional Exogenous Variables	National Exogenous Variables	Total Exogenous Variables
ALASKA Tuck	49	0	49
CALIFORNIA Burton Dyckman	340	17	357
CALIFORNIA Ratajczak	30	12	42
CALIFORNIA Roberts Wittels	77	11	88
SOUTHERN CALIFORNIA Moody Puffer	6	3	9
HAWAII Chau	35	3	38
MASSACHUSETTS Bell	1	8	9
OHIO L'Esperance Nestel Fromm	13	3	16
PUERTO RICO Dutta Su	17	5	22
PUERTO RICO Stahl	11	2	13

Source: Dean Chen [64].

for Stahl's model of Puerto Rico to the astonishing number of 357 exogenous variables for the Burton-Dyckman model of California.

A more detailed summary of these models is contained in the review articles that have been cited above. In view of the fact that the models are essentially elaborations on the representative regional model that has been discussed in some detail, no useful purpose would be served by introducing these detailed summaries here. To conclude this brief survey of regional models on an optimistic note, it is apparent that the results that have been obtained to date establish the fact that useful regional econometric models can be constructed. On a somewhat less optimistic note, very little experience has been gained in the use of these models so that the validation process for any one of these models is far from complete. Moreover, the problems of data availability impose very serious constraints on the detail that can be incorporated in these regional models. Much of the ingenuity of these models lies in the use of proxy variables in place of missing data. In spite of these difficulties, substantial progress has been made in regional econometric modeling over the past decade. And there is every reason to believe that future progress will be just as rapid.

IV. FORECASTING AND POLICY ANALYSIS: AN EXAMPLE

In an attempt to illustrate the use of econometric models in concrete terms, a recent forecast produced by the Michigan Quarterly Econometric Model of the U. S. Economy [92] is examined.

In addition, several policy experiments also based on this model are reviewed. These numerical results are intended to provide the policy-maker a realistic view of what can be expected from econometric models.

Forecasting With an Econometric Model. The Michigan Model [49] consists of 61 equations and is similar in outline to the model described in Section II. A recent set of forecasts produced by this model is shown in Tables V and VI. A careful examination of these tables reveals that the Michigan model prediction shows a reduction in the rate of growth of gross national product over the next two years. Nominal GNP is expected to grow at a rate of 11.8% over the 1972–73 period and taper off to 8.6% over the 1973–74 period. The forecast shows that real gross national product, i.e., nominal GNP corrected for price changes, is expected to grow at 6.2% over the 1972–73 period. A meager 2.4% increase is predicted over the following year.

At the same time that the economy is slowing down, the unemployment rate is expected to increase from 4.85% in 1973 to 5.35% in 1974. This increase in the aggregate unemployment rate is expected to be accompanied by an increase in the rate of inflation. Using the gross national product deflator to measure the price level, the projected rate of inflation is 5.2% for 1973 and 6.1% for 1974. Projections of various components of gross national product are given in these two tables.

Of perhaps more interest than these figures themselves is the way in which the forecast was constructed. In this connection the specification and estimation of the model is only the first step in the projection procedure. Since this is a true *ex ante* forecast, i.e., a forecast that is made before the fact, it is necessary to forecast a number of variables which are not explained by the model and which are therefore referred to as exogenous variables. The more important exogenous variables of the Michigan model are

- The Federal budget.
- State and local purchases of goods and services.
- Monetary policy variables such as the discount rate and the money supply.
- Farm prices, and
- Import prices.

Any error involved in the projected values of these and the other exogenous variables of the model will lead to an error in the forecast values of the endogenous variables or variables that are explained by the model. This is, of course, not the only source of error and it is possible for these errors to offset one another to some extent.

The likely error to which these projections are subject can be inferred from past forecasts. The usual method used to evaluate forecast accuracy is to derive what are called *ex post* forecasts. This involves the use of known values of the exogenous variables as inputs to the model. The corresponding values of the endogenous variables are then calculated and these solution values are compared with the actual values. Error statistics for selected endogenous variables are shown in Table VII. These error statistics were calculated for the period 1960.1–1969.4. The variables in this table are defined as follows:

GNP : Gross National Product, billions of 1958 dollars.

GNP$: Gross National Product, billions of current dollars.

u : Unemployment Rate, percent.

PGNP : Gross National Product Implicit Deflator; 1958 = 100.

These results can be used to determine the precision of the one-quarter forecasts that are obtained on the basis of perfect information about the exogenous variables. For example, if the forecast errors are normally distributed a 95% confidence interval (2-standard deviations) would be approximately ± 5.6 billion for real GNP. A 95% confidence interval for the unemployment rate would be ± .38. Using this

Table V Michigan Model Forecast

Gross National Product in Current Dollars
(Billions of Current Dollars, Seasonally Adjusted at Annual Rates)

	1973			1974				Calendar Years		% Change	
	II[1]	III	IV	I	II	III	IV	1973	1974	1972–73	1973–74
Gross National Product	1271.0	1309.81	1341.80	1368.15	1389.94	1412.47	1439.26	1291.28	1402.46	11.8	8.6
Personal Consumption Expenditures	795.1	815.55	834.18	851.59	866.12	881.38	898.72	806.07	874.45	11.0	8.5
Durable Goods	133.1	135.23	136.98	138.23	137.96	139.32	142.73	134.39	139.56	14.5	3.9
Automobiles and Parts	59.9	59.71	60.30	59.62	58.63	58.37	60.33	60.09	59.24	13.8	−1.4
Furniture and Household Equipment	54.5	56.57	57.34	58.90	59.25	60.50	61.64	55.52	60.07	15.4	8.2
Other Durables	18.8	18.95	19.34	19.71	20.07	20.45	20.77	18.78	20.25	14.0	7.8
Nondurable Goods	329.8	340.00	348.75	356.68	363.19	368.79	374.48	335.18	365.79	11.8	9.1
Services	332.2	340.33	348.45	356.67	364.97	373.27	381.51	336.47	369.11	8.8	9.7
Gross Private Domestic Investment	199.2	208.20	213.70	213.71	213.75	213.75	214.78	203.91	214.00	14.3	4.9
Nonresidential	134.4	139.79	145.87	150.83	154.81	157.64	159.69	137.75	155.74	16.5	13.1
Residential Structures	59.5	58.62	55.58	52.08	49.42	48.90	49.94	58.17	50.09	7.7	−13.9
Change in Business Inventories	5.3	9.79	12.25	10.81	9.52	7.21	5.14	7.99	8.17	—	—
Net Exports	1.7	2.36	2.52	3.55	4.68	5.94	6.97	1.64	5.28	—	—
Exports	95.1	98.76	102.50	106.39	110.44	114.63	118.98	96.51	112.61	31.4	16.7
Imports	93.4	96.40	99.98	102.84	105.76	108.69	112.01	94.88	107.33	21.6	13.1
Government Purchases of Goods and Services	275.0	283.70	291.40	299.30	305.40	311.40	318.80	279.65	308.72	9.7	10.4
Federal	106.5	108.20	109.30	113.00	114.10	115.50	117.90	107.37	115.12	2.8	7.2
National Defense	74.5	75.20	75.30	77.20	77.30	77.70	78.10	74.82	77.57	0.6	3.7
Other	32.0	33.00	34.00	35.80	36.80	37.80	39.80	32.55	37.55	8.2	15.4
State and Local	168.5	175.50	182.10	186.30	191.30	195.90	200.90	172.27	193.60	14.4	12.4
Gross National Product Deflator, (1958 = 100)	152.30	155.05	157.79	160.18	162.20	164.16	166.01	153.74	163.14	5.2	6.1
Aggregate Unemployment Rate (%)	4.92	4.68	4.77	4.96	5.23	5.50	5.70	4.85	5.35	—	—

[1] U. S. Department of Commerce, July 1973, preliminary.
Source: Hymans and Shapiro [92].

Table VI Michigan Model Forecast

Gross National Product in Constant Dollars
(Billions of 1958 Dollars, Seasonally Adjusted at Annual Rates)

	1973			1974				Calendar Years		% Change	
	II[1]	III	IV	I	II	III	IV	1973	1974	1972–73	1973–74
Gross National Product	834.6	844.77	850.38	854.13	856.94	860.44	866.99	839.76	859.62	6.2	2.4
Personal Consumption Expenditures	553.8	557.57	559.84	564.70	568.20	572.65	578.87	555.98	571.10	5.5	2.7
Durable Goods	116.6	117.47	117.52	117.99	117.16	117.42	119.05	117.14	117.91	12.6	0.7
Automobiles and Parts	53.0	52.47	52.07	51.39	50.46	49.80	50.61	52.92	50.57	12.6	—4.4
Furniture and Household Equipment	48.3	49.66	49.88	50.80	50.70	51.36	51.97	48.95	51.21	12.9	4.6
Other Durables	15.3	15.34	15.57	15.79	16.01	16.25	16.48	15.28	16.13	11.7	5.6
Nondurable Goods	228.1	228.57	228.51	230.48	232.46	234.30	236.53	228.50	233.44	3.4	2.2
Services	209.0	211.53	213.81	216.23	218.58	220.93	223.29	210.34	219.75	4.2	4.5
Gross Private Domestic Investment	131.0	134.71	136.25	134.14	132.03	129.72	128.23	133.02	131.03	8.3	—1.5
Nonresidential	91.8	93.62	95.82	97.48	98.36	98.57	98.42	93.11	98.21	11.2	5.5
Residential Structures	35.3	33.89	31.43	28.71	26.67	25.85	26.03	34.05	26.82	—1.4	—21.2
Change in Business Inventories	3.9	7.20	9.01	7.95	7.00	5.30	3.78	5.85	6.01	—	—
Net Exports	4.9	5.35	5.52	6.22	6.97	7.78	8.44	4.43	7.35	—	—
Exports	65.3	66.80	68.33	69.90	71.51	73.15	74.83	66.42	72.34	17.7	8.9
Imports	60.4	61.45	62.81	63.68	64.54	65.37	66.39	61.99	64.99	6.1	4.8
Government Purchases of Goods and Services	144.9	147.14	148.77	149.07	149.74	150.29	151.46	146.29	150.14	2.3	2.6
Gross National Product Deflator, (1958 = 100)	152.30	155.05	157.79	160.18	162.20	164.16	166.01	153.74	163.14	5.2	6.1
Aggregate Unemployment Rate (%)	4.92	4.68	4.77	4.96	5.23	5.50	5.70	4.85	5.35	—	—

[1] U. S. Department of Commerce, July 1973, preliminary.
Source: Hymans and Shapiro [92].

TABLE VII

Ex Post Forecast Error Statistics, The Michigan Model, 1960.1–1969.4

Variable	Root Mean Squared Error	Error Statistic Mean Absolute	Mean Absolute Percentage Error
GNP	2.808	2.287	.397
GNP$	2.941	2.370	.370
u	.190	.152	4.316
PGNP	.149	.129	.115

Source: Hymans and Shapiro [93].

historical experience as a guide, the most likely value of the unemployment rate for the third quarter of 1973 is 4.68%. A 95% confidence interval would be 4.68 ± .38 or 4.30 to 5.06. It would be surprising indeed if the actual unemployment rate were to fall outside this interval. This rare event is expected to happen no more than once in every twenty forecasts.

These calculations of the precision of the model forecasts are conservative in the sense that they do not take into account potential errors in the projection of the exogenous variables. As mentioned previously, actual *ex ante* forecasts are conditional on the exogenous variables. If fiscal or monetary policy turns out after the fact to be substantially different than was predicted, the forecast error of the endogenous variables is also likely to be much larger.

The entries in Table VII relate only to one-quarter forecasts. It is also of interest to examine how forecast accuracy deteriorates as the forecast horizon is increased. Error statistics for multiperiod predictions are shown in Table VIII for the same variables that are contained in Table VII. The point to note is the rate of deterioration of forecast accuracy. For example, four-quarter error in the unemployment rate forecast is double the one-period standard error. The errors in the GNP projection do not accumulate as rapidly but the deflator errors build up even more rapidly. These results are consistent with the qualification that usually accompanies an eight quarter forecast to the effect that the second year is illustrative only and indicative of likely broad tendencies.

Other types of error statistics can be calculated for models such as the Michigan Model. For example, the number of business-cycle turning points that are accurately predicted can be computed. The use of these and other *ex post* error statistics to validate and compare models is considered in the appendix to this chapter. Perhaps the most important point of this discussion is that the policy-maker should be aware that econometric models are subject to prediction error. He should insist that error statistics such as those described above should accompany the model and any forecasts that are presented. Only then will he have an idea of the range of error that is likely to be observed.

Policy Analysis With An Econometric Model. Alternative economic policies can be examined with the aid of an econometric model. Two policy changes that illustrate some of the dynamic properties of the Michigan Model are discussed in this section. The first involves a permanent increase of $1 billion in State and local government purchases of goods and services. The other involves the use of monetary policy to lower the treasury bill rate by 100 basis points.

The numerical results of the fiscal policy experiment are shown in Table IX. The new variable in this table, UM%, is the unemployment rate of males, 20 years of age and older. The

TABLE VIII

Root Mean Squared Errors for One to Four Quarter Forecasts Over 1961.1–1969.4

	Number of Quarters			
Variable	1	2	3	4
GNP	3.19	3.60	4.81	4.76
GNP$	3.43	4.18	5.66	5.87
u	.21	.27	.30	.42
PGNP	.14	.27	.24	.32

Source: Hymans and Shapiro [93].

TABLE IX

Fiscal Policy Simulation With the Michigan Model: Response to a One Billion (1958 Dollars) Increase In State and Local Government Purchases

Quarter	GNP	Response UM%	PGMP
1	.85	−.04	.02
2	1.64	−.10	.02
3	2.16	−.15	.02
4	2.59	−.20	.01
5	2.92	−.22	.02
6	3.17	−.24	.04
7	3.32	−.25	.05
8	3.39	−.25	.09
9	3.36	−.25	.14
10	3.22	−.23	.20
11	2.96	−.21	.27
12	2.59	−.18	.35
13	2.10	−.13	.43
14	1.45	−.08	.56
15	.65	−.02	.58
16	−.22	+.05	.63

Source: Hymans and Shapiro [93].

results show that the response to the increase in government spending reaches a peak some eight quarters after the policy is initiated and then begins to lose its impact. At the end of fifteen quarters, the income effects are negligible. The unemployment rate reflects this same pattern with a cumulative reduction in the unemployment rate until the eighth quarter followed by an increase. At the 16th quarter, the economy ends up with no change in real GNP, no change in the unemployment rate, but a higher price level than would have been the case without the increase in government spending. However, all during the period GNP has been higher and the unemployment rate lower than would have been the case. The price of this increase in real output is the inflation that would be experienced over the period.

The results of the monetary policy experiment are shown in Table X. The interesting point of this simulation from the point of policy design is the long lag in the response to monetary policy. Not until after the eighth quarter does the response begin to build. Indeed, it is almost three years before the increase in real output is significant. Just as the predictions described previously are subject to error, so to are these policy simulations. These simulations give only likely affects on the assumption that there are no other changes except the policy change. But since these impacts are based on a model with estimated coefficients, the results may be quite misleading. Unfortunately, it is very difficult to calculate tolerance intervals for these multipliers. The reason for this is that there is generally a large number of unknown parameters which must be estimated, as many as 100 or 200. It is just not feasible to vary these parameters both individually and jointly to determine the sensitivity of the solution to these parameter variations. Such sensitivity analyses, even on a small scale, would be extremely valuable in an evaluation of the usefulness of the model for policy analysis purposes. The policy-maker should routinely insist on such studies so that undue precision will not be attached to the results of policy simulations.

Concluding Remarks. This section has attempted to illustrate the use of econometric models in forecasting and

TABLE X

Monetary Policy Simulation With the Michigan Model: Response to a 100 Basis Point Reduction In the Treasury Bill Rate

Quarter	GNP
1	0
2	−.02
3	−.06
4	−.09
5	−.10
6	−.09
7	−.06
8	.03
9	.19
10	.40
11	.61
12	.99
13	1.74
14	2.66
15	3.34
16	3.60

Source: Hymans and Shapiro [93].

policy analysis. The general theme that has been emphasized is that the models that are available today are subject to error. In some cases, these errors are substantial. The policy-maker should insist on tolerance intervals for forecasts and on sensitivity studies for simulation experiments. Otherwise, he is likely to be misled as to the accuracy of the results generated by the model.

The scale of many of the models that are available may at first appear to be a deterrent to their use by policy-makers. However, with the spread of time-shared systems, it has become a relatively simple matter to use these models. Policy simulations of the variety described previously can be performed routinely with the software that is now available. It is possible to subscribe for a nominal fee to DRI, the Wharton Model, and others. Such a subscription typically includes quarterly forecasts and entitles the user to access the model for policy experimentation. Thus the use of even very complex models should become rather routine in the next several years. With such ready access, it is imperative that policy-makers understand the limitations of these models lest they be disappointed in the results.

References

Econometric Methods

1. M. Athans and G. C. Chow, "Introduction to Stochastic Control Theory and Economic Systems," *Annals of Social and Economic Measurement*, 1 (1972), 375–84.
2. G. C. Chow, "Optimal Stochastic Control of Linear Economic Systems," Journal of Money, Credit and Banking, 2 (1970), 291–302.
3. G. C. Chow, "Optimal Control of Linear Econometric Systems With Finite Time Horizon," *International Economic Review*, 13 (1972), 16–25.
4. C. F. Christ, *Econometric Models and Methods*, New York: Wiley, 1966.
5. P. J. Dhrymes, *et. al.*, "Criteria for Evaluation of Econometric Models," *Annals of Social and Economic Measurement*, 1 (1972), 291–324.
6. P. J. Dhrymes, *Econometrics: Statistical Foundations and Applications*, New York: Harper and Row, 1970.
7. J. W. Forrester, *Urban Dynamics*, Cambridge: The M.I.T. Press, 1969.
8. J. W. Forrester, *Industrial Dynamics*, Cambridge: The M.I.T. Press, 1961.
9. G. Fromm and L. R. Klein, "Solutions of the Complete System," in J. S. Duesenberry, G. Fromm, L. R. Klein and E. Kuh, eds., *The Brookings Model: Some Further Results*, Chicago: Rand McNally, 1969, 362–422.
10. A. S. Goldberger, *Econometric Theory*, New York: Wiley, 1964.
11. G. R. Green and L. R. Klein, "The Predictive Record of the Wharton Forecasts," *The Wharton Quarterly*, forthcoming.
12. Y. Haitovsky and N. Wallace, "A Study of Discretionary and Nondiscretionary Fiscal and Monetary Policies in the Context of Stochastic Macro-Econometric Model," in V. Zarnowitz, ed., *The Business Cycle Today*, National Bureau of Economic Research, 1972.
13. R. S. Holbrook, "Optimal Policy Choice Under a Nonlinear Constraint: An Iterative Application of Linear Techniques," Michigan Economic Research Paper, Series II, No. 1, Department of Economics, University of Michigan, Ann Arbor, 1973.
14. E. P. Howrey, "Selection and Evaluation of Econometric Models," presented at the Symposium on Computer Simulation Versus Analytical Solution for Business and Economic Models, Gothenburg, Sweden, 1972.
15. E. P. Howrey and H. H. Kelejian, "Computer Simulation Versus Analytical Solutions," in T. H. Naylor, ed., *The Design of Computer Simulation Experiments*, Duke University Press, 1969.
16. E. P. Howrey, L. R. Klein, and M. D. McCarthy, "Notes on Testing the Predictive Performance of Econometric Models," Discussion Paper No. 173, Wharton School, Department of Economics, University of Pennsylvania, Philadelphia, 1970.
17. J. Johnston, *Econometric Methods*, 2nd Edition, New York: McGraw-Hill, 1972.
18. L. R. Klein, *A Textbook of Econometrics*, Evanston: Row, Peterson, 1953.
19. L. R. Klein, *An Essay on the Theory of Economic Prediction*, Yrjo Jahnsson Lectures, Helsinki: Academic Book Store, 1968.
20. L. R. Klein, "Forecasting and Policy Evaluation Using Large Scale Econometric Models: The State of the Art," *Frontiers of Quantitative Economics*, M. D. Intriligator, ed., Amsterdam: North Holland, 1971.
21. J. Kmenta, *Elements of Econometrics*, New York: MacMillan, 1971.
22. T. H. Naylor, "Policy Simulation Experiments with Macroeconomic Models: The State of the Art," *Frontiers in Quantitative Economics*, M. D. Intriligator, ed., Amsterdam: North Holland, 1971.
23. R. S. Pindyck, "Optimal Stabilization Policies via Deterministic Control," *Annals of Social and Economic Measurement*, 1 (1972), 385–90.

24. J. B. Ramsey, "Tests for Specification Errors in Classical Least-Squares Regression Analysis," *Journal of the Royal Statistical Society*, Series B, 1969, 350–71.
25. H. O. Stekler, "Forecasting with Econometric Models: An Evaluation," *Econometrica*, 36 (1968), 437–63.
26. D. B. Suits, "Forecasting and Analysis With an Econometric Model," *American Economic Review*, 52 (1962), 104–32.
27. D. B. Suits, "Applied Econometric Forecasting and Policy Analysis," *Forecasting on a Scientific Basis*, Proceedings of an International Summer Institute, Curia, Portugal, September 1968.
28. H. Theil, *Principles of Econometrics*, New York: Wiley, 1971.
29. H. Theil and J. C. G. Boot, "The Final Form of Econometric Equation Systems," *Review of the International Statistical Institute*, 30 (1962), 136–52.
30. Wonnacott, Ronald J. and Wonnacott, Thomas H., *Econometrics*, John Wiley & Sons, Inc., New York: 1970.
31. V. Zarnowitz, "New Plans and Results of Research in Economic Forecasting," Fifty-First Annual Report, National Bureau of Economic Research, 1971, 53–70.

National Econometric Models

32. F. G. Adams and L. R. Klein, "Predicting Consumer Behavior," a chapter in *Human Behavior in Economic Affairs. Essays in Honor of George Katona*, edited by B. Strumpel, J. N. Morgan, and E. Zahn, North-Holland Publishing Co., Amsterdam: 1972.
33. L. C. Andersen and K. M. Carlson, "A Monetarist Model for Economic Stabilization," *Federal Reserve Bank of St. Louis Review*, 52 (1970), 7–25.
34. A. Ando, F. Modigliani and R. Rasche, "Appendix to Part I: Equations and Definitions of Variables for the FRB-MIT-PENN Model, November, 1969," *Econometric Models of Cyclical Behavior*, B. Hickman, ed., New York: Columbia University Press, 1972.
35. T. M. Brown, *Specification and Uses of Econometric Models*, London: MacMillan, 1970.
36. C. F. Christ, "Aggregate Econometric Models," *American Economic Review*, 46 (1956), 385–408.
37. R. L. Cooper, "The Predictive Performance of Quarterly Econometric Models of the United States," *Econometric Models of Cyclical Behavior*, B. Hickman, ed., New York: Columbia University Press, 1972.
38. J. Duesenberry, O. Eckstein and G. Fromm, "A Simulation of the U. S. Economy in Recession," *Econometrica*, 38 (October 1960), 749–809.
39. J. S. Duesenberry, G. Fromm, L. R. Klein, and E. Kuh (eds.), *The Brookings Model: Some Further Results*, Chicago: Rand McNally, 1969.
40. J. S. Duesenberry, G. Fromm, L. R. Klein, and E. Kuh (eds.), *The Brookings Quarterly Econometric Model of the United States*, Chicago: Rand McNally, 1965.
41. M. K. Evans, Y. Haitovsky and G. I. Treyz, "An Analysis of the Forecasting Properties of U. S. Econometric Models," *Econometric Models of Cyclical Behavior*, B. Hickman, ed., New York: Columbia University Press, 1972.
42. M. K. Evans and L. R. Klein, *The Wharton Econometric Forecasting Model, 2nd, Enlarged Edition*, Studies in Quantitative Economics, Volume 2, Economic Research Unit, University of Pennsylvania, 1968.
43. M. K. Evans, L. R. Klein, and M. Saito, "Short-Run Prediction and Long-Run Simulations of the Wharton Model, *Econometric Models of Cyclical Behavior*, B. Hickman, Ed., New York: Columbia University Press, 1972.
44. R. C. Fair, *A Short-Run Forecasting Model of the U.S. Economy*, Lexington: Heath, 1971.
45. G. Fromm, "Inventories, Business Cycles, and Economic Stabilization," in *Inventory Fluctuations and Economic Stabilization*, Part IV, Joint Economic Committee, U. S. Congress 87th Session, Washington: U. S. Government Printing Office, 1962, 37–133.
46. G. Fromm and L. R. Klein, "A Comparison of Eleven Econometric Models of the United States," presented at the American Economic Association meetings, Toronto, 1972.
47. G. Fromm and P. Taubman, *Policy Simulations with an Econometric Model*, Washington: The Brookings Institution, 1968.
48. G. R. Green, "Short- and Long-Term Simulations with the OBE Econometric Model," *Econometric Models of Cyclical Behavior*, B. Hickman, ed., New York: Columbia University Press, 1972.
49. S. Hymans and H. Shapiro, "The Michigan Quarterly Econometric Econometric," Proceedings of the Conference on the Economic Outlook for 1973, Ann Arbor: University of Michigan, 1973.
50. L. R. Klein, *Economic Fluctuations in the United States, 1921–41*, New York: Wiley, 1950.
51. L. R. Klein and A. S. Goldberger, *An Econometric Model of the United States: 1929–52*, Amsterdam: North Holland, 1955.
52. L. R. Klein and J. Popkin, "An Econometric Analysis of the Post-War Relationship between Inventory Fluctuations and Changes in Aggregate Economic Activity," in *Inventory Fluctuations and Economic Stabilization*, Part III, Joint Economic Committee, U. S. Congress 87th Session, Washington: U. S. Government Printing Office, 1961, 69–89.
53. M. Liebenberg, A. A. Hirsch, and J. Popkin, "A Quarterly Econometric Model of the United States," *Survey of Current Business*, 46 (1966), 13–39.
54. T. C. Liu, "An Exploratory Quarterly Model of Effective Demand in the Postwar U. S. Economy," *Econometrica*, 31 (July, 1963), 301–48.
55. M. D. McCarthy, *The Wharton Quarterly Econometric Forecasting Model Mark III*, Studies in Quantitative Economics,

Volume 6, Economics Research Unit, University of Pennsylvania, 1972.

56. M. Nerlove, "A Tabular Survey of Macro-Econometric Models," *International Economic Review*, 7 (1966), 127–75.

57. J. Tingergen, *Business Cycles in the United States of America, 1919–1932*, Geneva: League of Nations, 1939.

Regional Econometric Models

58. F. W. Bell, "An Econometric Forecasting Model for a Region," *Journal of Regional Science*, 7 (1967), 109–127.

59. C. Blanco, "The Determinants of Interstate Population Movements," *Journal of Regional Science*, 5 (1963), 77–84.

60. C. Blanco, "Prospective Unemployment and Interstate Population Movements," *Review of Economics and Statistics*, 46 (1964), 221–222.

61. G. H. Borts, "The Equalization of Returns and Regional Economic Growth," *American Economic Review*, 50 (1960), 319–347.

62. R. P. Burton and J. W. Dyckman, *A Quarterly Economic Forecasting Model for the State of California*, Center for Planning and Development, Research Institute of Urban and Regional Development, University of California, Berkeley, 1965.

63. L. C. Chau, "An Econometric Model of Hawaii," University of Hawaii, 1970, mimeographed.

64. Dean Chen, "A Tabular Survey of Selected Regional Econometric Models," Working Paper No. 11, Federal Reserve Bank of San Francisco, July 1972.

65. R. T. Crow, *An Econometric Model of the Northeast Corridor of the United States*, Prepared for Northeast Corridor Transportation Project, U. S. Department of Transportation, March, 1969.

66. Dutta, M. and V. Su, *An Econometric Model of Puerto Rico*, mimeographed, 1968.

67. M. J. Emerson, "Regional Economic Projection Techniques," *Kansas Business Review*, 20, Number 3 (March 1967).

68. N. J. Glickman, "An Econometric Forecasting Model for the Philadelphia Region," *Journal of Regional Science*, 11 (1971), 15–32.

69. N. J. Glickman, "The Structure of a Large-Scale Regional Econometric Model: Towards the Simulation of Regional Development," *Discussion Paper No. 235*, University of Pennsylvania.

70. N. J. Glickman, "Regional Econometric Models: The Japanese and American Experience," paper presented at the Second Pacific Conference of the Regional Science Association, Tokyo, Japan, August, 1971.

71. R. Goldman, *The Structure of the Santa Clara County Model*, (County of Santa Clara: Planning Department), mimeographed, 1966.

72. R. J. Green, *A Long Range Econometric Forecasting Model for Illinois*, (University of Illinois, Bureau of Economic and Business Research), 1967.

73. W. L. Hansen and C. M. Tiebout, "An Intersectoral Flow Analysis of the California Economy," *Review of Economics and Statistics*, 45 (1963), 409–18.

74. Yokio Kaneko, "An Econometric Approach to Annual Forecast on Regional Economy by Local Government," *Regional Science Association Pacific Meetings*, 1965.

75. John W. Kendrick and C. Milton Jaycox, "The Concept and Estimation of Gross State Product," *Southern Economic Journal*, 32 (1965), 153–168.

76. L. R. Klein, "The Specification of Regional Econometric Models," *Papers of the Regional Science Association*, 23 (1969).

77. L. R. Klein, "A Model of Japanese Growth, 1878–1937," *Econometrica*, 29 (1961), 277–292.

78. W. L. L'Esperence, G. Nestel, And D. Fromm, "Gross State Product and an Econometric Model of a State," *American Statistical Association Journal*, 64 (1969), 787–807.

79. H. T. Moody and F. W. Puffer, "A Gross Regional Product Approach to Regional Modelbuilding," *Western Economic Journal*, 7 (1969), 391–402.

80. H. T. Moody, F. M. Puffer and P. M. Williams, "An Eight Region Model of the United States Economy," *Growth and Change*, 1 (1970), 20–26.

81. H. T. Moody and R. M. Williams, "Forecasting Gross Regional Product for Southern California," *California Management Review*, 1 (1967), 83–92.

82. F. Moore and J. Petersen, "Regional Analysis: An Interindustry Model of Utah," *Review of Economics and Statistics*, 37 (1955), 368–80.

83. F. W. Puffer and R. M. Williams, "An Econometric Model for Southern California and Forecasts for 1967," *California Management Review*, 3 (1967), 89–95.

84. D. Ratajczak, "A Method for Estimating Gross State Product," in R. M. Williams, S. Rabin and D. Ratajczak, "The Use of Anticipations Surveys in Forecasting for the Southern California Economy," *CIPET Conference*, Madrid, September 1969.

85. D. Ratajczak, "Simulation of the Economic Impact of State Witholding Upon the California Economy," *Proceedings 1972 Summer Computer Simulation Conference*, San Diego, 1217–1226.

86. Research Seminar in Quantitative Economics, *Econometric Model of Michigan*, Ann Arbor, Michigan: University of Michigan Press, 1965.

87. B. F. Roberts and G. Wittels, "An Econometric Model of California," Paper presented at the *Winter Meetings, Econometric Society*, December 1971.

88. J. E. Stahl, "An Application of a Klein Growth Model to Puerto Rico, 1947–61," *Economic Development and Cultural Change*, 13 (July 1965), 463–71.

89. H. Thomassen, "A Growth Model for a State," *Southern Economic Journal*, 24 (1957), 123–139.

90. B. Tuck, *An Aggregate Income Model of a Semi-Autonomous Economy*, Federal Field Committee for Development Planning in Alaska, Fairbanks: Alaska: 1967.

Miscellaneous

91. J. W. Forrester, *World Dynamics*, Cambridge: Wright Allan Press, 1971.
92. S. H. Hymans and H. T. Shapiro, "The Economic Outlook at Mid-Year," Ann Arbor: The University of Michigan, 1973.
93. ———, "The Michigan Quarterly Econometric Model of the U. S. Economy," Ann Arbor: The University of Michigan, 1973.
94. T. R. Lakshmanan and Fu-Chen Lo, "A Regional Model for the Assessment of Effects of Air Pollution Abatement," Environment and Planning, 4 (1972), 73–97.
95. D. H. Meadows, et. al., The Limits to Growth (A report for the Club of Rome's project on the Predicament of Mankind), New York: Universe Books, 1972.
96. H. deVries, "A Critical Assessment of the MIT World Models," presented at the Symposium, "Computer Simulation Versus Analytical Solutions for Business and Economic Models," Gothenburg, 1972.

Appendix

The purpose of this appendix is to provide a brief technical introduction to the construction and use of econometric models. In the first section of this appendix, model construction is considered. The interrelated steps of specification, estimation, and validation are described in some detail. Various simultaneous equation estimation methods are reviewed and several validation tests are summarized. The second section outlines the use of econometric models in forecasting and policy analysis.

Econometric Models

An econometric model was previously defined as an empirical economic model. There are three essential steps in the construction of an econometric model. First, the relationships of the model must be specified explicitly in a form amenable to statistical analysis. Second, the unknown parameters must be estimated from data on the variables contained in the model. Finally, the model must be validated to determine the extent to which the model is consistent with known properties of the system that is being modeled. These three aspects of model construction: specification, estimation, and validation, are discussed in this section.

Specification. The specification of the equations of the model is perhaps the most critical step in the construction of an econometric model. The reason for this is that the statistical procedures employed in estimation and hypothesis testing are predicated on the assumption that the specification is correct. An incorrect specification can lead to incorrect or misleading inferences being drawn from the model.

In the specification stage of an econometric study the basic functional relationships are set out. Returning to the example of the Introduction, the demand function is written as

$$D = D(p, q, y, t) \qquad (A\text{-}1)$$

where the variables D, p, q, y, and t are as defined previously. Writing the demand function in this general form is not without content. For even in this general form, the function indicates which variables are being entertained as direct determinants of the quantity demanded per unit of time. In this context, only four variables are included in the demand function.

In addition to the selection of variables, it is necessary at the specification stage to distinguish between *endogenous* and *exogenous* variables. Endogenous variables are those variables that are to be determined by the model. Exogenous variables are those variables determined outside the model which have a bearing on the determination of the endogenous variables of the model. The set of exogenous variables can be further subdivided into these variables that can be controlled or partially controlled by the policymaker and those which cannot be so controlled. In the commodity market example, the tax rate t can presumably be controlled by the policymaker whereas the price of competing commodities, q, cannot be directly controlled.

The number of variables included in the model depends to some extent on the purpose for which the model was constructed and on the nature of the process being modeled. The division of the variables between endogenous and exogenous similarly depends on these same considerations. For example, if the policymaker is specifically concerned with the development in a single commodity market, it may not be too great a distortion to assume that income, y, is exogenous. If the commodity market were the market for onions, this would be more appropriate than if it were the automobile market. In either case, the feedback from market developments to the level of national income would not be significant. If the single market is the only consideration, then the price of competing goods might also be taken as exogenous. However, if the policymaker is concerned with the agricultural support program, a single-market approach might not be appropriate since developments in one market may have important repercussions in other markets. In this case, it would not be appropriate to regard the prices of other goods as exogenous.

This may be summarized by noting that a *systems* approach is necessary at the specification stage of econometric modeling. All the important variables must be included in the system and all of the important interdependencies should be taken into account. The specification of the system is partially a matter of the nature of the problem and partly determined by the interest of the policymaker.

The formulation of the general system relationships is only part of the specification process. It is necessary, in addition, to decide on the functional forms to be employed. Here, statistical theory imposes certain constraints on the types of functional forms that can be employed. In general it is very difficult to deal with equations which are not linear in the unknown parameters of the system. Thus a linear approximation to (A-1), given by (1′) in the text, or a log-linear approximation to (1), given by

$$\ln D = \alpha_0 - \alpha_1 \ln(p + t) + \alpha_2 \ln q + \alpha_3 \ln q \qquad \text{(A-2)}$$

are linear in the unknown parameters α_0, α_1, α_2, and α_3. However, an equation of the form

$$D = \alpha_0 (p + t)^{-\alpha_1} q^{\alpha_2} + \alpha_3 y \qquad \text{(A-3)}$$

is not linear in the parameters and would lead to severe problems in parameter estimation.

Ideally, the theory on which the model is based would indicate the functional forms to be used. However, economic theory is frequently not sufficiently precise to indicate the variables that should be included in the equation, not to mention the functional form. This is particularly true when it comes to dynamic systems. It is generally true in commodity models, for example, that the quantity demanded per month responds with a distributed lag to price changes. That is, because of inertia in buying habits and lack of information about potential substitutes, consumers adjust with a lag to price changes. This might lead to the dynamic linear specification

$$D_j = \alpha_0 - \alpha_1 (p + t)_j + \alpha_2 q_j + \alpha_3 y_j + \alpha_4 D_{j-1} \qquad \text{(A-4)}$$

for example. However, this is only one of a variety of lag patterns that could have been introduced and economic theory often provides almost no guidance in this matter.

A final aspect of the specification phase of an econometric study involves the introduction of the stochastic properties of the model. No matter how complex the model, it is extremely unlikely that the model will correspond exactly to the real system that is modelled. This inexactness is recognized by introducing an error term with certain properties in the equations of the system. For example, the equation (A-4) becomes

$$D_j = \alpha_0 - \alpha_1 (p + t)_j + \alpha_2 q_j + \alpha_3 y_j + \alpha_4 D_{j-1} + u_j \qquad \text{(A-5)}$$

where u_j is a random variable with a zero mean and finite variance σ^2. In addition, it is often assumed that the error term exhibits no serial correlation so that u_t and u_{t-j} are uncorrelated for $j \neq 0$. A specific density function such as the normal density is also frequently assumed for the disturbance term but this is not always the case.

By way of summarizing the specification stage of econometric modeling, the *standard linear econometric model* is introduced. The model contains n jointly dependent, endogenous variables denoted by the vector y and m predetermined variables denoted by the vector z. These predetermined variables include both exogenous variables and lagged values of the endogenous variables. The model is written as

$$B y + \Gamma z = u \qquad \text{(A-6)}$$

where B is a matrix of dimension $n \times n$, Γ is a matrix of dimension $n \times m$, and u is a vector of n disturbances, one for each equation. The point to note is that there is one equation for each endogenous variable, otherwise the model would not be complete. Economic theory is used wherever possible to select the variables for each equation. Thus some of the elements of the coefficient matrices B and Γ will be zero. The problem now is to estimate the non-zero elements of these coefficient matrices.

Estimation. The unique feature of the linear econometric model (A-6) is the fact that the n endogenous variables in the vector y are simultaneously determined. Econometric theory has developed methods to estimate the parameters of the matrices B and Γ which take this simultaneity into account. Indeed, a number of estimation procedures have been developed and it is necessary to choose from among a set of estimators. The choice turns on several considerations. In some instances the form of the model will be sufficient to determine the appropriate estimation procedure. More often than not, however, the estimation procedure that is used will be a compromise between the ideal estimation procedure and a procedure that is feasible from a computational point of view. The following discussion attempts to acquaint the reader with several alternative estimation procedures and some of the criteria that can be used to choose among the estimators.

The simplest estimation procedure both conceptually and from the point of view of implementation is the method of least squares applied to each equation. If β_i denotes the ith row of B and γ_i denotes the ith row of Γ, the least squares estimator is obtained by minimizing

$$S_i = \sum_{j=1}^{N} (\beta_i y_j + \gamma_i z_j)^2 \qquad \text{(A-7)}$$

with respect to the elements of β_i and γ_i. The sum of squared errors, S_i, is defined over the sample of N observations on y and z. The entire set of parameter estimates, obtained by repeating this procedure for $i = 1, 2, \ldots, n$, are referred to as ordinary least squares (OLS) estimates.

The OLS estimates do not, in general, have very attractive properties. In particular, the estimates of the parameter β_i are not consistent. That is, as the sample size increases without limit, the estimate of β_i does not generally converge to β_i. This rather serious problem does not arise in one particular case; that is, if the model is *fully recursive*. A model is said to be recursive if the matrix of structural coefficients B is triangular and the covariance matrix of the disturbance vector u is diagonal. What this means is that y_1 is jointly determined by $y_2, y_3, \ldots, y_n$; y_2 is jointly determined by $y_3, y_4, \ldots, y_n$; and so on. In addition the disturbance in the first equation is uncorrelated with the disturbances in all other equations. In this special case, ordinary least squares yields best linear unbiased estimates of the parameters. That is, the expected value of the least square estimates of the parameters are the actual values and the estimates have a smaller variance than any other estimator that is a linear function of the observations on y. Except in this special case, ordinary least squares is not recommended.

It might be useful to point out that

the problem with the use of ordinary least squares arises in the attempt to estimate the structural coefficient in the matrix B. One way around this problem is to focus attention on the *reduced form* of the model which is given by

$$y = \pi z + v \qquad \text{(A-8)}$$

where $\pi = -B^{-1}\Gamma$ and $v = B^{-1}u$. Each equation is now a function of only predetermined variables and the simultaneous equation problem does not arise. The application of ordinary least squares to each reduced-form equation yields an estimate $\hat{\pi}$ of the coefficient matrix π. Provided the predetermined variables are uncorrelated with the disturbances, this estimate has the desirable properties of a least-squares estimator.

There are several potential problems with this approach, however. First, the full vector z which now enters each equation of the reduced form system (11) may have a large number of elements. Indeed, it is not unusual in large models for the dimension of z to exceed the number of observations that are available to estimate $\hat{\pi}$. In this case the reduced-form estimator cannot be used. A second problem is that the reduced form estimator does not use any of the restrictions on the coefficients matrices B and Γ to estimate π. A more efficient estimation procedure would use this information. Finally, the coefficients of B and Γ are typically of direct interest to the policymaker. Only in the special case in which the model is exactly identified is it possible to unscramble the relationship $\hat{\pi} = -B^{-1}\Gamma$ to obtain estimates of B and Γ. For these reasons, reduced form estimation is not generally advocated.

Perhaps the most popular estimator that avoids the pitfalls of ordinary least squares but retains much of the simplicity of the least squares procedure is Two-Stage Least Squares (TSLS). The essential features of TSLS can be illustrated by a consideration of the first equation of the system which is written as

$$y_{1t} + \sum_{j=2}^{n} \beta_{1j} y_j + \sum_{j=1}^{m} \gamma_{1j} z_{jt} = u_{1t} \qquad \text{(A-9)}$$

Typically, a number of the coefficients β_{1j} and γ_{1j} will be zero. Let $\beta_{1.}'$ and $\gamma_{1.}'$ denote the nonzero coefficients. Then (A-9) can be written as

$$y_{1t} + \sum_{j=2}^{n'} \beta'_{1j} y_{jt} + \sum_{j=1}^{m'} \gamma'_{1j} z_{jt} = u_{1t} \qquad \text{(A-10)}$$

with an appropriate reordering of the endogenous and predetermined variables, if necessary. In the first stage of the two-stage estimator, the endogenous variables included in this equation, y_2, y_3, ... , $y_{n'}$, are regressed on the predetermined variables as in the reduced form equations (A-8). Reduced-form predictions of y_2, y_3, ... , $y_{n'}$ are then obtained from

$$\hat{y}_t = \hat{\pi} z_t \qquad \text{(A-11)}$$

for the sample period $t = 1, 2, \ldots, N$. The parameters in (A-10) are then obtained by minimizing the sum of squared errors

$$S_1 = \sum_{t=1}^{N} \Big(y_{1t} + \sum_{j=2}^{n'} \beta'_{1j} \hat{y}_{jt} + \sum_{j=1}^{m'} \gamma'_{1j} z_{jt}\Big)^2 \qquad \text{(A-12)}$$

with respect to the unknown parameters.

This estimation procedure is not without its difficulties. If the number of predetermined variables exceeds the number of observations, the first-stage, reduced form estimator cannot be computed. In this case a procedure that is employed is to regress y_j not on the full set of variables z, but rather on the first m'' principle components of this set. This circumvents the problem but introduces the need to choose m'' and the first-stage estimates and hence $\hat{y}$ and thus the second-stage estimates may be sensitive to the value assigned to m''.

Other estimation procedures are often discussed but seldom used in practice. These alternative estimators will only be mentioned here. The two-stage estimator uses some of the information about the complete system but not as much as it might. In particular, it takes into account the fact that $y_1, y_2, \ldots, y_n$ are jointly determined but it does not use the fact that the deviations u_1 may be correlated with the deviations u_j $(j = 2, 3, \ldots, n)$. If the covariance matrix of the disturbances process is not diagonal, this should be taken into account in the estimation of the parameters. The estimation procedure referred to as *three-stage least squares* accomplishes this. The estimation procedure has not been employed very frequently in the estimation of parameters of econometric models so it will not be discussed further.

The two- and three-stage procedures make use of system information but do not incorporate potential knowledge about the distribution of the disturbance vector. The method of *maximum likelihood* can be applied to derive parameter estimation procedures if the distribution of the disturbance vector is known. It is generally necessary to assume that the error process is normal for lack of further information. The computational problems become rather prohibitive in all but the simplest cases. Hence, maximum likelihood procedures are not used very frequently in the estimation of national and regional econometric models.

By way of summarizing this discussion of parameter estimation, the practical choice among estimation procedures usually reduces to ordinary least squares applied to the structural equations, ordinary least squares applied to the reduced form or two-stage least squares. There is no simple rule-of-thumb that can be given to select the best estimator from this group. The preference from the point of view of the theory of estimation is for the two-stage estimator *if* the model is correctly specified. This is a severe qualification. It means that the investigator should be convinced that the constraints imposed on the system in the form of zero coefficients for some variables are correct. Otherwise more may be lost by imposing incorrect constraints than by disregarding the constraints altogether and applying ordinary least squares to the original system, equation by equation.

Validation. An important aspect of any econometric study is the validation of the model that has been constructed. Validation generally refers to the process by which the validity of the model is established. This process generally involves successive confirmation of the model as a useful decision-making device. A sequence of successful applications of the model provides the basis for the determination of the circumstances in which the model contributes significantly to the decision-making process. In this sense, the validation of a model is a continuing process.

In view of the fact that the specification and estimation of an econometric model are frequently an interdependent process, the model that eventually emerges from this stage of the investigation may be sample-dependent to an undesirable degree. What typically happens in model construction is that a preliminary model is specified and the parameters are estimated. Some of the estimated parameters turn out to be insignificant so those variables are dropped from the model. Inevitably other variables are added to the model and the dynamic, distributed-lag relationships are empirically determined to a large extent. Thus the empirical model that emerges may depend on the peculiarities of the sample data to an unwarranted degree. It is desirable to test for this possibility, if at all possible, before using the model in a decision-making context.

At the present time, there is no universally accepted set of tests of validity of an econometric model. However, the work of Howrey [14] and Dhrymes, *et al.* [5] suggest that it is useful to view the problem of validation in the following way. Three types of tests are of particular interest in connection with validation. These are

- Tests of data conformity,
- Comparisons of alternative models, and
- User-oriented criteria of success.

These tests are discussed in detail in Howrey [14] and are merely summarized here to give the reader an idea of what is involved.

Tests of data conformity attempt to determine the extent to which the data are consistent with the empirical model. Some of the tests that have been introduced in this connection can be based on the same data that are used in parameter estimation. However, more powerful tests are generally obtained if a different set of data is used to validate the model. A type of test that uses post-sample data, i.e., data beyond the sample period that was used in the estimation of the parameters, is the test for structural stability. Returning to the first equation of the system (A-6), the empirical equation is

$$y_{1t} + \sum_{j=2}^{n} b_{1j} y_{jt} + \sum_{j=1}^{m} c_{1j} z_{jt} = 0 \tag{A-13}$$

where b_{1j} is the estimate of β_{1j} and c_{1j} is the estimate of γ_{1j}. It is assumed that these parameter estimates are based on a sample of N observations ($t = 1, 2, \ldots, N$). Now suppose that τ additional observations for $t = N + 1, N + 2, \ldots, N + \tau$ are available. To test for structural stability, the probability that these additional observations were generated by the empirical equation (A-13) is calculated. If this probability is small, say less than .05, then the model is rejected. Otherwise the data are consistent with the model and the model is retained.

The model might fail this test for a variety of reasons. The coefficients might have changed between $t = 1, 2, \ldots, N$ and $t = N + 1, N + 2, \ldots, N + \tau$. The coefficients b_{1j} or c_{1j} may have erroneously large or small values. Relevant variables may have been omitted from the equation. An incorrect functional form may have been employed. There are clearly a number of potential errors in the model. On the other hand, a significant departure from (14) is required to reject the model. Thus it is not clear in general how stringent such a test might be. As indicated by Cooper [37], the test does discriminate against several aggregate econometric models that have been published.

A battery of tests based on the sample data used to estimate the parameters of the model has been developed by Ramsey [24]. These tests are concerned with testing for correct functional form, testing for omitted variables, and testing for correlation of the regressors with the disturbance process. Although these tests are not now performed routinely, they provide useful information which is relevant to the validation of a model. If a model passes all these tests it is retained, otherwise it is rejected.

The tests described thus far are concerned with the evaluation of a single model and are aimed at providing an answer to the single question "Is the model consistent with the data?" A second phase of the validation process involves comparisons of the model with alternative descriptions of reality. This is an exercise in hypothesis testing in which the model under investigation is treated as the null or maintained hypothesis while alternative models are considered as potential candidate models. The motivation for tests of this type derive from a consideration of the prediction accuracy of a model. If the predictions of a complicated econometric model are not at least as good as those of a simple extrapolative model, the econometric model is rejected. Formal tests of this sort turn out to be very complicated unless one model can be written as a submodel of the other. If this is true, then the two hypotheses are *nested* and classical statistical methodology can be used to test the hypotheses. If the hypotheses are not nested, a relatively simple test procedure is not available. In this case several approximate tests are discussed in Howrey [14].

A final set of tests of the validity of a model involves an evaluation of the ability of the model to aid the policy-maker. Zarnowitz [31] has considered the problem of validation from a user point of view and finds that it is not easy to implement. The conditions that must be met include the following:

- the errors must be identifiable
- the preferences of decision makers must be known
- the constraints under which the decisionmaker operates must be known
- the cost of using the model must be known.

If all these are available then the "success" of the model can be measured by the reduction in the cost of incorrect decisions which are avoided by the use of the model. This user-oriented approach to model validation is to some extent subjective and depends heavily on the specific objectives of each user. Hence, there are no published examples of this type of validation. Nevertheless, it is an important element in the evaluation of an econometric model. As models are used in the decision-making process, it is expected that their use will be under continuous review.

Forecasting, Analysis and Control

One of the primary reasons for the construction of an econometric model is to use the model as an aid in economic forecasting. In addition to this, however, models are also extremely useful in the examination of alternative economic policies. Going one step further, a model can be used to aid in the design of optimal policies. These three areas constitute the important potential uses of econometric models and will be considered in some detail in this section.

Forecasting With An Econometric Model. In this discussion of the use of an econometric model in a forecasting context, it is convenient to rewrite the reduced-form system of equations (A-8) as

$$y_t = B_1 y_{t-1} + B_2 y_{t-2} + \ldots + B_r y_{t-r} + C\,x_t + v_t \qquad \text{(A-14)}$$

This result is obtained from the reduced-form equation by separating the lagged endogenous vectors y_{t-1}, y_{t-2}, . . . , y_{t-r} from the exogenous variables x_t in the vector z_t of predetermined variables. The matrices B_1, B_2, . . . , B_r, and C are obtained from $\hat{\pi}$ and denote matrices of estimated coefficients. The vector v_t of disturbances is assumed to be serially uncorrelated.

Several types of forecasts can be generated using this equation. It is useful to distinguish among these different types of forecasts especially in connection with the analysis of forecast error and in connection with the comparison of alternative models. A first distinction that is useful is between *ex ante* and *ex post* forecasts. An *ex ante* forecast refers to a forecast that is made before the values of the exogenous variables are known. If $y_t{}^1$ denotes the predicted value of y_t one period in advance, then it follows from (A-14) that

$$y_t{}^1 = \sum_{j=1}^{r} B_j y_{t-j} + C\,x_t{}^1 \qquad \text{(A-15)}$$

is the one-period, *ex ante* forecast of the vector y_t. The one-period, *ex post* forecast of y_t is given by

$$y_t{}^1 = \sum_{j=1}^{r} B_j y_{t-j} + C\,x_t \qquad \text{(A-16)}$$

The difference between (A-16) and (A-15) is that in (A-15) the exogenous variables are forecast whereas in (A-16) the actual values are used.

Ex post forecasts are particularly useful for an investigation of the validity of the model. The *ex post* forecast error depends on the disturbance v_t and the errors in the estimates of the coefficient matrices. The *ex ante* forecasts include an additional component that results from errors in the predictions of the exogenous variables. This *ex ante* forecast error is useful for determining the potential range of error of unconditional forecasts whereas the *ex post* forecast error indicates the range of error of conditional forecasts.

In addition to the one-period forecasts, $y_t{}^1$, multi-period forecasts can be generated using (A-14). For example, the two-period, *ex ante* forecast is

$$y_t^2 = B_1 y^1_{t-1} + B_2 y_{t-2} + \dots + B_r y_{t-r} + C x_t^2$$

In general, a ν-period forecast is given by

$$y_t^\nu = \sum_{k=1}^{\nu-1} B_k y_{t-k}^{\nu-k} + \sum_{k=\nu}^{r} B_k y_{t-k} + C x_t^\nu \quad \text{(A-17)}$$

An additional source of error arises in the multiperiod forecast since the lagged endogenous vectors $y_{t-1}, y_{t-2}, \dots, y_{t-\nu+1}$ must be forecast in advance. This dynamic forecast gives rise to the accumulation of forecast error which is of interest in the comparison of models. A model that exhibits a slow error buildup is generally preferable to one with a rapid error accumulation.

This discussion of forecasting with an econometric model has been couched in terms of the linear model (A-14). However, many current econometric models contain some nonlinearities so that this discussion is not completely relevant. Fortunately, it is not difficult to modify the results given above to include the case of nonlinear models. In particular, it is assumed that the reduced form can be written as

$$y_t = F(y_{t-1}, y_{t-1}, \dots, y_{t-r}, x_t, v_t) \quad \text{(A-18)}$$

where $F[\cdot]$ denotes a vector of functions. By analogy with the linear model, a one-period, *ex ante* prediction is defined by

$$y_t^1 = F(y_{t-1}, y_{t-2}, \dots, y_{t-r}, x_t^1, 0) \quad \text{(A-19)}$$

As Howrey and Kelejian pointed out [15], this is not an unbiased forecast since, in general, the expectation of a function of a random variable is not equal to the function of the expectation of the random variable. More recently, Haitovsky and Wallace [12] have found that the bias in (A-19) is indeed statistically significant for at least one macroeconomic model.

An unbiased one-period, *ex ante* forecast could be obtained by numerical methods. By averaging over a number of replications of

$$\hat{y}_t^1 = F(y_{t-1}, y_{t-2}, \dots, y_{t-r}, \hat{x}_t^1, \hat{v}_t) \quad \text{(A-20)}$$

an unbiased forecast is obtained. Each replication in (A-20) involves an evaluation of the function F(·) with a randomly generated value of $\hat{x}^1_t$ and $\hat{v}_t$. The situation is even more complicated in the multi-period prediction case. In order to obtain an unbiased forecast, replications of

$$\hat{y}_t^\nu = F(y^{\nu-1}_{t-1}, y^{\nu-2}_{t-2}, \dots, \hat{y}^1_{t-\nu+1}, y_{t-\nu}, \dots, y_{t-r}, \hat{x}_t^\nu, \hat{v}_t) \quad \text{(A-21)}$$

must be obtained and averaged. Although no empirical work has been conducted on this problem, it is almost certain that the unbiased multi-period prediction would deviate significantly from the corresponding biased ν-period prediction given by

$$y_t^\nu = F(y^{\nu-1}_{t-1}, \dots, y^1_{t-\nu+1}, y_{t-\nu}, \dots, y_{t-r}, x_t^\nu, 0) \quad \text{(A-22)}$$

For any particular non-linear model, there is a need to investigate the impact of nonlinearities on the predictions generated by the model.

Policy Analysis With An Econometric Model. Policy analysis with a linear econometric model can be illustrated by returning to the reduced-form system (A-14). The vector x_t of exogenous variables is partitioned into a subvector of control variables, denoted by g_t, and a subvector h_t of independent, or uncontrolled, variables. The matrix C is partitioned conformably so that the reduced form system can be written as

$$y_t = \sum_{j=1}^{r} B_j y_{t-j} + P g_t + Q h_t + v_t \quad \text{(A-23)}$$

Alternative policies are characterized by assigning different values to the elements of g_t. The properties of alternative policies can be determined from this reduced-form system.

The impact multiplier associated with a change in one of the elements of the vector of control variables is defined by

$$\partial y_t / \partial g_{jt} = P_{.j} \qquad \text{(A-24)}$$

which is the jth column of the matrix *P*. The immediate impact of a unit change in g_{jt} on y_{kt} is thus P_{kj}. The matrix *P* is the matrix of *impact multipliers* that are used to compare the immediate effects of alternative policy changes.

This same system of equations can be used to calculate the *dynamic multipliers* associated with a policy change. The term dynamic multiplier refers to the time path of the response of the vector y_t to a change in g_{jt}. That is, the response in period $t + k$ to a unit change in the control variable g_j in period t is

$$\begin{aligned} \partial y_{t+k} / \partial g_{jt} &= B_1 \partial y_{t+k-1} / \partial g_{jt} \\ &\quad + B_2 \partial y_{t+k-2} / \partial g_{jt} + \ldots \\ &\quad + B_k \partial y_t / g_{jt} \\ &= [B_1{}^k + B_2{}^{k-1} + \ldots \\ &\quad + B_k] P_{\cdot j} \qquad \text{(A-25)} \end{aligned}$$

for $k = 1, 2, \ldots$.

The impact and dynamic multipliers for a nonlinear model can be defined in an analogous way. For example, the vector of impact multipliers corresponding to a change in g_j is given by

$$\begin{aligned} \partial y_t / \partial g_{jt} = \partial F(&y_{t-1}, \ldots, y_{t-r}, \\ &g_t, h_t, 0) / \partial g_j \qquad \text{(A-26)} \end{aligned}$$

where $F(\cdot)$ is the vector of functions of the reduced-form (A-18). It is again necessary to point out that, in general, this is a biased estimate of the impact multiplier. It is, nevertheless, the multiplier that is usually calculated. An equally important point to notice about this multiplier is that, in general, it will not be a constant but will depend on the values assumed by the predetermined variables. This added complication is one of the features that makes it rather difficult to work with nonlinear models.

The dynamic multiplier $\partial y_{t+k} / \partial g_{jt}$ is given by

$$\begin{aligned} \partial y_{t+k} / \partial g_{jt} = [&\partial F(\cdot) / \partial y_{t+k-1}] \\ &\partial y_{t+k-1} / \partial g_{jt} \\ &+ [\partial F(\cdot) / \partial y_{t+k-2}] \partial y_{t+k-2} / \partial g_{jt} \\ &+ \ldots + [\partial F(\cdot) / \partial y_t] \partial y_t / \partial g_{jt} \\ &\qquad \text{(A-27)} \end{aligned}$$

Once again, the sequence of multipliers defined in this way is biased. Moreover, the sequence is a function of the time paths of the variables of the system.

It is useful to consider briefly the way in which the dynamic multipliers defined in (A-27) are calculated in practice. First, a control solution $y_t{}^c$ for $t = r + 1, \ r + 2, \ldots, \ r + T$ corresponding to the initial conditions, $y_r{}^c$, $y^c{}_{r-1}, \ldots, y_1{}^c$ and the control path $g_t{}^c$, $h_t{}^c$ $t = r + 1, r + 2, \ldots, r + T$ is obtained from (19) with $v_t = 0$. This control solution provides the basis for comparison of alternative policies. A new solution corresponding to $g_t{}'$ is obtained starting from the same initial conditions and using the same values for $h_t{}^c$. The difference between the new solution $y_t{}'$ and the control solution $y_t{}^c$ is the response path to the policy change given by $g_t{}' - g_t{}^c$. Generally, an iterative method is used to solve the nonlinear system. Some of the algorithms that have proved to be useful in this connection are described by Fromm and Klein in [9].

Optimal Economic Policy. In addition to the analysis of alternative *ad hoc* policies as described above, it is possible to use an econometric model in the design of optimal economic policy. In order to do this, it is necessary to introduce explicitly an objective function which compares alternative outcomes of the system. The objective or welfare function is written as

$$\begin{aligned} W = W(&y_1, y_2, \ldots, y_T; \\ &g_1, g_2, \ldots, g_T) \qquad \text{(A-28)} \end{aligned}$$

to indicate that the policymaker is concerned with the values of the endogenous variables y_t as well as the values of the control variables g_t over the planning horizon $t = 1, 2, \ldots, T$. The values assumed by the exogenous variables h_t may also be of interest to the policymaker, but since they are beyond his control, there is no need to include them explicitly in the objective function.

The economic policy problem is to maximize the welfare function subject to the constraints imposed by the economic system characterized by (A-18). In this general form, this is a non-linear programming problem with stochastic constraints which is a rather intractable problem. Hence, it is necessary to simplify the problem in order to obtain a problem that it is feasible to solve.

A simplification that makes the problem somewhat more tractable is the introduction of a quadratic, separable welfare function of the form

$$W = \sum_{t=1}^{T} (y_t - y_t^*)'\mathbf{K}_t^1(y_t - y_t^*) + \sum_{t=1}^{T} (g_t - g_t^*)'\mathbf{K}_t^2(g_t - g_t^*) \qquad \text{(A-29)}$$

In this formulation, y_t^* and g_t^* denote optimal values of the endogenous and control variables and $\mathbf{K}_t^1$ and $\mathbf{K}_t^2$ are symmetric positive semidefinite matrices which assign costs to deviations of y_t and g_t from the optimal values. For example, if $\mathbf{K}_t^1$ is a diagonal matrix with diagonal elements k_t^{jj}, then the contribution of y_t to the welfare function is

$$\sum_{j=1}^{n} (y_{jt} - y_{jt}^*)^2 k_t^{jj}.$$

If the diagonal elements were all equal, then the squared deviations $(y_{jt} - y_{jt}^*)^2$ would all have equal weight in the objective function.

If this welfare function is combined with the linear model (A-14), a solution can be obtained without a great deal of difficulty. This type of economic control problem has been discussed by Chow [2], Pyndick [23], and others. If the econometric model is not linear, the usual procedure is to employ a linear approximation in the neighborhood of the control solution. An iterative solution process based on such a linear approximation is discussed in a recent paper by Holbrook [13].

It is readily apparent that one of the major difficulties with this approach to the design of economic policy is the specification of the objective function. To the extent that the policies derived from the optimization procedure are sensitive to small changes in the objective function, the policymaker may be reluctant to implement the policies. If optimal economic policies are this sensitive, an awareness of the sensitivity seems desirable.

Optimal economic control with an econometric model is still in its infancy. Much remains to be done before optimal control methods can be applied with any degree of confidence. Even so, it is desirable for policymakers to be aware of these developments so that they will be prepared to describe their preferences in the coherent and rigorous manner that is required by the control-theoretic approach to policymaking.

☆ U.S. GOVERNMENT PRINTING OFFICE: 1974 O—535-365